10/19/31
DIALOG(R)File 347:JAPIO

04818188 **Image available**
DISK CACHE DEVICE

PUB. NO.: 07-110788 [JP 7110788 A]
PUBLISHED: April 25, 1995 (19950425)
INVENTOR(s): YORIMITSU KEIICHI
IWATANI SAWAO
APPLICANT(s): FUJITSU LTD [000522] (A Japanese Company or Corporation), JP
(Japan)
APPL. NO.: 05-256216 [JP 93256216]
FILED: October 14, 1993 (19931014)
INTL CLASS: [6] G06F-012/08
JAPIO CLASS: 45.2 (INFORMATION PROCESSING -- Memory Units)

ABSTRACT

PURPOSE: To perform an optimized disk cache operation suited to the successive application form of a user and a disk array by selectively setting a reliability considering mode, a performance considering mode or an average mode.

CONSTITUTION: Thus device is provided with a pair of disk control means 12 for reading and writing the data of a disk device 30 for access requests from a host device 10, a cache storage means 48 for storing a part at the data stored in the disk device 30, a cache control means 46 and a shared bus tar connecting the pair of the disk control means 12 and transmitting and receiving the information of the access requests or the like and the data. In this case, the cache control means 46 is provided with a reliability considering mode cache control means 72, a state considering mode cache control means 74-1 and an average mode cache control means 6-1. Then, cache control corresponding to an operating mode set by an operating mode setting means 70 at the time of starting the device is performed.

EG
2 of 2

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報**（Ａ）　(11)特許出願公開番号

特開平7－110788

(43)公開日　平成７年(1995)４月25日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　12/08	３２０	7608－5Ｂ		

審査請求　未請求　請求項の数36　ＯＬ　（全 42 頁）

(21)出願番号　特願平5－256216

(22)出願日　平成５年(1993)10月14日

(71)出願人　000005223
富士通株式会社
神奈川県川崎市中原区上小田中1015番地

(72)発明者　依光　圭一
神奈川県川崎市中原区上小田中1015番地
富士通株式会社内

(72)発明者　岩谷　沢男
神奈川県川崎市中原区上小田中1015番地
富士通株式会社内

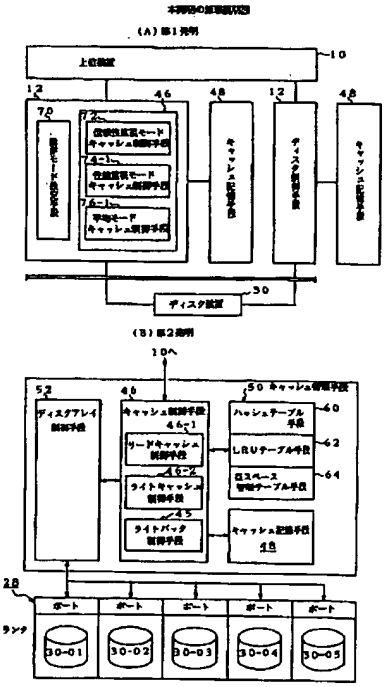
(74)代理人　弁理士　竹内　進　（外１名）

(54)【発明の名称】　ディスクキャッシュ装置

(57)【要約】

【目的】ユーザの運用形態とディスクアレイに適合して最適化されたディスクキャッシュ動作を提供する。

【構成】２組のコントローラに各々キャッシュ機構を設け、各系統に同一データを格納する信頼性重視モード、各系統ごとに独立にデータを格納する性能重視モード、両者を折衷した平均モードのいずれかで動作可能とする。ＲＡＩＤ０，１，３，５て動作するディスクアレイを対象としたキャッシュ処理を効率的に行なう。ＲＡＩＤ５の動作でもディスクリードを必要とせずにパリティを計算できるＲＡＩＤ３的なアクセスを可能とするようにキャッシャブロック上に有効データを準備する。



EG
1/92

【特許請求の範囲】

【請求項１】ディスクキャッシュ装置に於いて、
上位装置（１０）からのアクセス要求に対しディスク装置（３０）のデータ読み書きを制御する一対のディスク制御手段（１２）と、
前記一対のディスク制御手段（１２）の各々に設けられ、前記ディスク装置（３０）に格納されたデータの一部を記憶するキャッシュ記憶手段（４８）と、
前記一対のディスク制御手段（１２）の各々に設けられ、前記上位装置（１０）からのアクセス要求に対し予め準備された信頼性重視モードキャッシュ制御手段（７２）、性能重視モードキャッシュ制御手段（７４）、又は平均モードキャッシュ制御手段（７６）にいずれか１つにより前記キャッシュ記憶手段（４８）のデータを読み書きする一対のキャッシュ制御手段（４６）と、
前記一対のディスク制御手段（１２）の間を結合して情報およびデータの送受を行う共用バス手段（１８）と、
装置の立ち上がり時に前記一対のキャッシュ制御手段（４６）に対し予め準備された信頼性重視モード、性能重視モードまたは平均モードにいずれか１つの動作モードを設定する動作モード設定手段（７０）と、を備えたことを特徴とするディスクキャッシュ装置。

【請求項２】請求項１記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記信頼性重視モードキャッシュ制御手段（７２）の各々は、前記一対のディスク制御手段（１２）に設けたキャッシュ記憶手段（４８）の各々に、同一のデータを記憶することを特徴とするディスクキャッシュ装置。

【請求項３】請求項２記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記信頼性重視モードキャッシュ制御手段（７２）の各々は、前記上位装置（１０）からのリード要求に対し自己のキャッシュ記憶手段（４８）に該当データがなかった場合は、前記ディスク装置（３０）から該当データを読出して自己のキャッシュ記憶手段（４８）にステージングして前記上位装置（１０）にデータを転送すると共に、前記共通バス手段（１８）を介して他のディスク制御手段（１２）に転送して他方のキャッシュ記憶手段（４８）に同じデータをステージングさせることを特徴とするディスクキャッシュ装置。

【請求項４】請求項２記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記信頼性重視モードキャッシュ制御手段（７２）の各々は、前記上位装置（１０）からのライト要求に対し自己のキャッシュ記憶手段（４８）にデータを書込むと共に、前記共通バス手段（１８）を介して他のディスク制御手段（１２）に転送して他方のキャッシュ記憶手段（４８）にも同じデータを書込ませることを特徴とするディスクキャッシュ装置。

【請求項５】請求項４記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記信頼性重視モードキャッシュ制御手段（７２）の各々は、前記上位装置（１０）からのライト要求の処理終了後に予め定めた書戻し条件が成立した際に、自己のキャッシュ記憶手段（４８）からディスク装置（３０）への書込みが済んでいないデータを抽出して前記ディスク装置（３０）に書き戻すと共に、前記共用バス手段（１８）を介して該書戻しデータのキャッシュ無効化を他方のディスク制御手段（１２）に指示することを特徴とするディスクキャッシュ装置。

【請求項６】請求項１記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記性能重視モードキャッシュ制御手段（７４）の各々は、前記一対のキャッシュ記憶手段（４８）の各々に、自己にアクセス要求のあったデータのみを個別に記憶することを特徴とするディスクキャッシュ装置。

【請求項７】請求項６記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記性能重視モードキャッシュ制御手段（７４）の各々は、前記上位装置（１０）からのリード要求に対し自己のキャッシュ記憶手段（４８）に該当データがなかった場合は、他方のディスク制御手段（１２）に対し前記共用バス手段（１８）を介して処理を依頼し、他のディスク制御手段（１２）からのデータなしの応答を受けた際に前記ディスク手段から該当データを読出して自己のキャッシュ記憶手段（４８）にステージングして前記上位装置（１０）にデータを転送することを特徴とするディスクキャッシュ装置。

【請求項８】請求項６記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記性能重視モードキャッシュ制御手段（７４）の各々は、前記上位装置（１０）からのライト要求に対し自己のキャッシュ記憶手段（４８）にのみデータを書込むことを特徴とするディスクキャッシュ装置。

【請求項９】請求項８記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記性能重視モードキャッシュ制御手段（７４）の各々は、前記上位装置（１０）からのライト要求の処理終了後に予め定めた書戻し条件が成立した際に、自己のキャッシュ記憶手段（４８）からディスク装置（３０）への書込みが済んでいないデータを抽出して前記ディスク装置（３０）に書き戻すことを特徴とするディスクキャッシュ装置。

【請求項１０】請求項１記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記平均モードキャッシュ制御手段（７６）の各々は、前記上位装置（１０）からのリード要求に伴う前記ディスク装置（３０）からのステージングを自己のキャッシュ記憶手段（４８）に対してのみ行い、上位装置（１０）からのライト要求に伴うデータの書込みは、前

記一対のディスク制御手段（１２）のキャッシュ記憶手段（４８）の各々に行って同一データを記憶することを特徴とするディスクキャッシュ装置。

【請求項１１】請求項１０記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記平均モードキャッシュ制御手段（７６）の各々は、前記上位装置（１０）からのリード要求に対し自己のキャッシュ記憶手段（４８）に該当データがなかった場合は、他方のディスク制御手段（１２）に対し前記共用バス手段（１８）を介して処理を依頼し、該依頼後に前記ディスク手段から該当データを読出して自己のキャッシュ記憶手段（４８）にステージングすることを特徴とするディスクキャッシュ装置。

【請求項１２】請求項１１記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記平均モードキャッシュ制御手段（７６）の各々は、他方のディスク制御手段（１２）への処理依頼に対しデータなしの応答を受けていた際には、前記ディスク手段から該当データを読出して自己のキャッシュ記憶手段（４８）にステージングした後に上位装置（１０）に転送することを特徴とするディスクキャッシュ装置。

【請求項１３】請求項１０記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記平均モードキャッシュ制御手段（７６）の各々は、前記上位装置（１０）からのライト要求に対し自己のキャッシュ記憶手段（４８）にデータを書込むと共に、前記共通バス手段（１８）を介して他のディスク制御手段（１２）に転送して他方のキャッシュ記憶手段（４８）にも同じデータを書込ませることを特徴とするディスクキャッシュ装置。

【請求項１４】請求項１３記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けた前記平均モードキャッシュ制御手段（７６）の各々は、前記上位装置（１０）からのライト要求の処理終了後に予め定めた書戻し条件が成立した際に、自己のキャッシュ記憶手段（４８）からディスク装置（３０）への書込みが済んでいないデータを抽出して前記ディスク装置（３０）に書き戻すと共に、前記共用バス手段（１８）を介して該書戻しデータのキャッシュ無効化を他方のディスク制御手段（１２）に指示することを特徴とするディスクキャッシュ装置。

【請求項１５】請求項１乃至１４記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けたキャッシュ記憶手段（４８）の各々を、不揮発性メモリ手段としたことを特徴とするディスクキャッシュ装置。

【請求項１６】請求項１乃至５記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（１２）に設けたキャッシュ記憶手段（４８）を揮発性メモリ手段としたことを特徴とするディスクキャッシュ装置。

【請求項１７】請求項１乃至１６記載のディスクキャッシュ装置に於いて、前記ディスク装置（３０）をディスクアレイ装置としたことを特徴とするディスクキャッシュ装置。

【請求項１８】ディスクキャッシュ装置に於いて、

内蔵した記憶媒体に対しデータを読み書きする複数のディスク装置（３０）を設けたディスクアレイ手段（２８）と、

前記ディスクアレイ手段（２８）に格納されたデータの一部を記憶するキャッシュ記憶手段（４８）と、

キャッシュ登録内容を示すハッシュテーブル手段（６０）および最新に使用されたキャッシュブロックを先頭位置とするように有効データを含むキャッシュブロックを登録純にリンクするＬＲＵテーブル手段（６２）に基づいて前記キャッシュ記憶手段（４８）の記憶状態を管理するキャッシュ管理手段（５０）と、

前記上位装置（１０）からリード要求を受けた際に前記キャッシュ管理手段（５０）を参照して該当データが存在する場合は前記キャッシュ記憶手段（４８）から読出して前記上位装置（１０）に転送し、該当データが存在しない場合は、前記ディスクアレイ手段（２８）からステージングした後に前記上位装置（１０）に転送するリードキャッシュ制御手段（４６－１）と、

前記上位装置（１０）からライト要求を受けた際に、前記キャッシュ管理手段（５０）に管理情報を登録すると共に該管理情報に従って前記キャッシュ記憶手段（４８）にデータを書込むライトキャッシュ制御手段（４６－２）と、

予め定めたライトバック条件が成立した際に、前記キャッシュ記憶手段（４８）から前記ディスクアレイ手段（２８）への記憶が済んでいないデータを抽出して書戻すライトバック制御手段（４５）と、

前記リードキャッシュ制御手段（４６－１）および前記ライトキャッシュ制御手段（４６－２）による前記ディスクアレイ手段（２８）のアクセス時に、予め設定されたディスクアレイの動作モードに従って１又は複数のディスク装置（３０）のアクセスを制御するディスクアレイ制御手段（５２）と、を備えたことを特徴とするディスクキャッシュ装置。

【請求項１９】請求項１８記載のディスクキャッシュ装置に於いて、前記ディスクアレイ手段（２８）は、並列的に配置された入出力ポートの各々にディスク装置（３０）を接続して１ランクを構成し、該ランク構成のディスク装置（３０）群を複数ランク設けたことを特徴とするディスクキャッシュ装置。

【請求項２０】請求項１９記載のディスクキャッシュ装置に於いて、前記ディスクアレイ制御手段（５２）は、ＲＡＩＤ０に対応した第１動作モードの設定時は、特定のランクの中のディスク装置（３０）を前記上位装置（１０）で扱う論理デバイスに１対１に割当て、個々の

ディスク装置（３０）に読み書きを行うことを特徴とするディスクキャッシュ装置。

【請求項２１】請求項１９記載のディスクキャッシュ装置に於いて、前記ディスクアレイ制御手段（５２）は、ＲＡＩＤ１に対応した第２動作モードの設定時は、特定のランクの中の２台のディスク装置（３０）を１組として前記上位装置（１０）で扱う論理デバイスに１対１に割当て、２つのディスク装置（３０）に同一データを書込むと共にいずれか一方のディスク装置（３０）からデータを読出すことを特徴とするディスクキャッシュ装置。

【請求項２２】請求項１９記載のディスクキャッシュ装置に於いて、前記ディスクアレイ制御手段（５２）は、ＲＡＩＤ３に対応した第３動作モードの設定時は、特定のランクを構成するｎ台のディスク装置（３０）をデータ用とパリティ用とに固定的に割当て、書込要求時に、書込データをビット単位又はバイト単位に（ｎ－１）分割すると共に分割単位ごとにパリティデータを計算し、前記（ｎ－１）分割したデータおよび前記パリティデータを前記ｎ台のディスク装置（３０）に並列的に書込むことを特徴とするディスクキャッシュ装置。

【請求項２３】請求項１９記載のディスクキャッシュ装置に於いて、前記ディスクアレイ制御手段（５２）は、ＲＡＩＤ５に対応した第４動作モードの設定時は、特定のランクを構成するｎ台のディスク装置（３０）を１組としてアクセスごとにパリティ位置が変化するように割当て、書込要求時に、書込データを少なくともディスク装置（３０）のセクタ単位に分割すると共に該書込データの書込先ディスク装置（３０）およびパリティ用ディクス装置から旧データおよび旧パリティデータを読出して新パリティを計算し、前記書込データ及び新パリティデータを前記書込先ディスク装置（３０）およびパリティ用ディクス装置に並列的に書込むことを特徴とするディスクキャッシュ装置。

【請求項２４】請求項１８記載のディスクキャッシュ装置に於いて、前記キャッシュ管理手段（５０）は、前記キャッシュ記憶手段（４８）の空き状態を管理する空スペース管理テーブル手段（６４）を備え、前記ライトキャッシュ制御手段（４６－１）は前記上位装置（１０）に基づくキャッシュ書込処理の完了報告時に前記空スペース管理テーブル手段（６０）を参照して空スペースが予め定めた閾値以下の場合、ディスク書込済みのデータは無効化して前記キャッシュ記憶手段（４８）から削除し、且つディスク書込みが済んでいないデータは前記ライトバック制御手段（４５）による処理対象に組込むことを特徴とするディスクキャッシュ装置。

【請求項２５】請求項２４記載のディスクキャッシュ装置に於いて、前記空スペース管理テーブル手段（６４）は、前記キャッシュ記憶手段（４８）上の連続するキャッシュブロックの空数をエントリとして先頭ブロックの位置情報を格納したことを特徴とするディスクキャッシュ装置。

【請求項２６】請求項１８記載のディスクキャッシュ装置に於いて、前記リードキャッシュ制御手段（４６－１）は、前記上位装置（１０）からの要求データが前記キャッシュ記憶手段（４８）の格納単位であるキャッシュブロック上に存在しない時、該キャッシュブロック上に存在する有効データの状態に応じて前記ディスクアレイ手段（２８）から新たなデータのステージングを行って連続する一つのデータ領域を生成することを特徴とするディスクキャッシュ装置。

【請求項２７】請求項２６記載のディスクキャッシュ装置に於いて、前記リードキャッシュ制御手段（４６－１）は、ステージング終了後に、前記要求データに続く有効データを前記ディスクアレイ手段（２８）から前記キャッシュブロック上にプリフェッチすることを特徴とするディスクキャッシュ装置。

【請求項２８】請求項２７記載のディスクキャッシュ装置に於いて、ＲＡＩＤ０に対応する第１動作モードの設定時又はＲＡＩＤ１に対応する第３動作モードの設定時に、前記リードキャッシュ制御手段（４６－１）は、ステージング終了後に前記要求データに続く予め設定した量の有効データを前記ディスクアレイ手段（２８）からプリフェッチすることを特徴とするディスクキャッシュ装置。

【請求項２９】請求項２７記載のディスクキャッシュ装置に於いて、ＲＡＩＤ３に対応する第３動作モードの設定時又はＲＡＩＤ５に対応する第４動作モードの設定時に、前記リードキャッシュ制御手段（４６－１）は、ステージング終了後に前記要求データに続く前記キャッシュブロックの最後までの有効データを前記ディスクアレイ手段（２８）からプリフェッチすることを特徴とするディスクキャッシュ装置。

【請求項３０】請求項２７記載のディスクキャッシュ装置に於いて、ＲＡＩＤ５に対応する第４動作モードの設定時に、前記リードキャッシュ制御手段（４６－１）は、ステージング終了後に前記要求データを含むパリティグループの最後までの有効データを前記ディスクアレイ手段（２８）からプリフェッチすることを特徴とするディスクキャッシュ装置。

【請求項３１】請求項１８記載のディスクキャッシュ装置に於いて、ＲＡＩＤ５に対応する第４動作モードの設定時に、前記リードキャッシュ制御手段（４６－１）は、オプション指定により前記キャッシュ記憶手段（４８）にパリティデータを含めてステージングし、前記ライトバック手段（４５）は、前記オプション指定を判別した際に前記キャッシュ記憶手段（４８）のパリティデータを用いて新たなパリティを計算することを特徴とするディスクキャッシュ装置。

【請求項３２】請求項１８記載のディスクキャッシュ装

置に於いて、前記リードキャッシュ制御手段（４６－１）は、前記上位装置（１０）から大量シーケンシャルデータのリードが要求された場合、前記ＬＲＵテーブル手段（６２）の終端からの一定量となる既存のキャッシュブロックの一部を無効化すると共に、前記大量シーケンシャルデータを前記ＬＲＵテーブル手段（６２）の先頭に登録せずに終端側に登録することを特徴とするディスクキャッシュ装置。

【請求項３３】請求項１８記載のディスクキャッシュ装置に於いて、ＲＡＩＤ５に相当する第４動作モードの設定時に、前記ライトバック制御手段（４５）は、前記キャッシュブロックの中のライトバックの対象となった有効データをパリティグループ単位に抽出してパリティデータを算出した後に、抽出した有効データおよびパリティデータを前記ディスクアレイ手段（２８）の複数のディスク装置（３０）に並列的に書込むことを特徴とするディスクキャッシュ装置。

【請求項３４】請求項１８記載のディスクキャッシュ装置に於いて、ＲＡＩＤ５に相当する第４動作モードの設定時に、前記ライトバック制御手段（４５）は、前記キャッシュブロックの中のライトバックの対象となった有効データのパリティグループに対する不足数が所定数以下の場合、該不足データを対応するディスク装置（３０）から読出した後にパリティグループ分の有効データおよび計算したパリティデータを前記ディスクアレイ手段（２８）の複数のディスク装置（３０）に並列的に書込むことを特徴とするディスクキャッシュ装置。

【請求項３５】請求項３３又は３４記載のディスクキャッシュ装置に於いて、前記ライトバック手段（４５）は、所定のライトバック条件が成立した際に、前記ＬＲＵテーブル手段（６２）内の終端から所定数のキャッシュブロックをライトバック対象に組込むと共に、該対象範囲に含まれるキャッシュブロックをデータが連続するように並び替えることを特徴とするディスクキャッシュ装置。

【請求項３６】請求項１８記載のディスクキャッシュ装置に於いて、更に、装置の使用中に前記上位装置（１０）からのアクセス要求で使用されるデータブロックサイズを統計情報として収集し、該統計情報の平均値を次の装置立ち上げ時の前記キャッシュ記憶手段（４８）のキャッシュブロックサイズとして自動設定する統計処理手段を設けたことを特徴とするディスクキャッシュ装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、ホストコンピュータからの指示により磁気ディスクサブシステムのリードとライトを行うキャッシュメモリを備えたディスクキャッシュ装置に関し、更に、磁気ディスクサブシステムにディスクアレイを用いたディスクキャッシュ装置に関する。

【０００２】近年、コンピュータ本体の性能の向上に伴い、コンピュータ本体のデータ格納媒体である磁気ディスク装置を用いた入出力サブシステムの性能向上も望まれている。このため磁気ディスク装置よりもビットコストは高いが、高速アクセスが可能な半導体メモリをキャッシュメモリとして用いたディスクキャッシュ装置が登場した。

【０００３】すなわち、磁気ディスクサブシステムの一部のデータをキャッシュメモリ上に置き、キャッシュメモリをリードまたはライトすることでアクセスを高速化するという手法である。初期のディスクキャッシュ装置には、揮発性のメモリ素子が使用されており、電源障害等の発生でデータ消失に至る可能性があり、リード処理はキャッシュヒットになれば高速に処理できる。

【０００４】しかし、ライト処理は磁気ディスク装置とキャッシュメモリの両方の同時更新を実行しなければならないため、高速化できなかった。その後、不揮発性のメモリ素子を使用したディスクキャッシュ装置が出るに及び、ライト処理においてはライトデータの更新をキャッシュ上のみで行い、ライト処理の完了後にディスク装置に書き込むライトバック処理が可能となった。

【０００５】このライトバック処理の実現によりライト処理の高速化が行われた。しかし、メモリを不揮発性を維持するためのバッテリ容量とコストの問題等で不揮発性メモリと揮発性メモリを混在させ、コストと性能の最適化を図った装置が多い。しかし、不揮発性メモリと揮発性メモリを混在させたディスクキャッシュ装置では、ライト要求がリード要求よりも圧倒的に多いユーザ環境では、ライトバック処理を行う余裕がなく、ユーザに適応出来ないことが多く、より一般的で高速性と信頼性重視のある装置が望まれている。

【０００６】更に近年、ディスクアレイ装置を構成する複数の磁気ディスクを組合せ、ＲＡＩＤとして知られた動作形態をとるものが普及しだしている。これらＲＡＩＤ型ディスクアレイ装置に対してもディスクキャッシュ装置を適用し、最適化を図ることが望まれている。

【０００７】

【従来の技術】図４１は揮発性キャッシュメモリのみで構成した従来のディスクキャッシュ装置を示す。ホストコンピュータ１００からチャネル１０４－１，１０４－２を介してアクセスされる２台のコントローラ１０２－１，１０２－２が接続される。コントローラ１０２－１，１０２は、下位のドライブインタフェースによりアダプタ１０８－１，１０８－２を介してドライブとして磁気ディスク装置１１０－１，１１０－２，１１０－３を接続する。

【０００８】またコントローラ１０２－１，１０２－０２は、下位のキャッシュインタフェースを介して共用される不揮発性のメモリ素子を用いたキャッシュメモリ１０６を接続している。図４２はキャッシュメモリに要求

データが存在した場合の動作を示す。ホストコンピュータ１００から例えばコントローラ１０２－１にリード要求①があると、コントローラ１０２－１はキャッシュメモリ１０６に要求データが存在するか否かをハッシュテーブルを参照して調べるヒット判定②を行う。

【０００９】すなわちキャッシュヒットの判定は、通常、磁気ディスク装置のアクセス単位であるシリンダ、トラック、ブロック等のハッシュパラメータからハッシュアドレス（ハッシュエントリ）を計算し、計算したハッシュアドレスで指定される要求データがハッシュテーブルに登録されているか否かをチェックする。要求データの存在を示すヒット判定が得られると、キャッシュメモリ１０６からデータ③を読出してホストコンピュータ１００に転送する。

【００１０】一方、キャッシュメモリ１０６は要領的に制限があることから、ＬＲＵテーブル（Least Recent Use Table）を用いたキャッシュ管理が行われる。ＬＲＵテーブルは最新に使用されたキャッシュブロックをテーブルの先頭位置に登録し、最も古いキャッシュブロックがテーブルの最後に位置することになる。ＬＲＵテーブルの登録が一杯になると古いブロックから破棄し新たなブロックを登録する。また一定期間アクセスがないキャッシュブロックは、キャッシュメモリの使用効率を高めるため破棄する。このキャシュブロックの登録を破棄してそのスペースを確保する処理をキャッシュ無効化という。キャッシュ無効化はハッシュテーブルの登録を削除することで実現できる。

【００１１】図４３はキャッシュメモリに要求データが存在しなかった場合の動作を示す。ホストコンピュータ１００からのリード要求①に対するヒット判定②でデータが存在しないミスヒットになると、磁気ディスク装置１１０から要求データをリードしてキャッシュメモリ１０６に書込むステージング③を実行する。このステージング③は、磁気ディスク装置１１０の目的シリンダにヘッドを位置付けるシーク動作とシーク完了後のヘッド選択に基づく目的トラックへのオントラック動作等の機械的動作を伴うことから、キャッシュメモリ１０６のアクセスに比べ時間がかかる。

【００１２】コントローラ１０２－１はキャッシュメモリ１０６のステージング③が完了すると、このデータ④をホストコンピュータ１００に転送し、ハッシュテーブルにデータが存在すること（ヒット状態）を登録する。図４４はライト要求に対する図４１の装置の処理動作を示す。ホストコンピュータ１００からライト指示①をコントローラ１０２－１が受けると、キャッシュメモリ１０６と磁気ディスク装置１１０の両方にデータを書込むライトスルー②を行う。

【００１３】これはキャッシュメモリ１０６が揮発性であり、電源障害等でデータの消失，データ化け等発生する可能性があり、信頼性に対する考慮からコントローラ１０２はデータをキャッシュメモリ１０６に書込むと同時に磁気ディスク装置１１０にも書込む。ライトスルーされたデータは次回のリードでは必ずキャッシュヒットとなる。即ち、ライト処理に要する時間は、キャッシュなしの場合と同じ時間が必要でキャッシュ導入の効果はない。

【００１４】ここで、図４１の従来装置のライト処理は、磁気ディスク装置とキャッシュメモリの両方を同時更新を実行しなければならないため、高速化できなかった。そこで図４５に示す揮発性キャッシュメモリと不揮発性キャッシュメモリを使用した装置が現われた。図４５において、ホストコンピュータ２００にチャネル２０４－１，２０４－２を介して接続された２台のコントローラ２０２－１，２０２－２は、下位のデバイスインタフェース側を全てのモジュール間での通信等を可能とする共有バス２０６を介して接続する。

【００１５】共用バス２０６にはコントローラ２０２－１，２０２－２で共用される揮発性キャッシュメモリ２０６と不揮発性メモリ２１０が設けられる。更に共用バス２０６にコントローラ２１２－１，２１２－２とアダプタ２１４－１，２１４－２を介して複数の磁気ディスク装置２１６－１，２１６－２が接続される。ここで各系統ごとに複数のコントローラを設けて各々に固有の処理機能に持たせるマルチコントローラ構成とすることで、コントローラ単体の場合に比べて負荷を軽減させ、全体のスループットを向上させている。

【００１６】揮発性キャッシュメモリ２０８は図４１の従来装置で使用したものと同様である。不揮発性キャッシュメモリ２１０は、ディスク装置に対するデータ書込を高速化するために設けられ、揮発性キャッシュメモリ２０８に対する書込データ２１８と同じデータが書込まれる。不揮発性キャッシュメモリ２１０に書込まれたデータは、ライト処理の完了後にライトバックという処理で磁気ディスク装置にに書込まれる。

【００１７】図４６はリード要求がミスヒットとなった時の図４５の装置のステージングを示す。４６において、ホストコンピュータ２００からリード要求①を受けると、コントローラ２０２－１は不揮発性メモリ２０８のヒット判定②を行う。ミスヒットの場合は、磁気ディスク装置２１６から要求テータをリードするステージング③を行い、キャッシュ管理テーブルに登録した後にホントコンピュータ２００にデータ転送④を行う。

【００１８】図４７は図４５の装置のライト処理を示す。４６において、ホストコンピュータ２００からライト指示①があると、コントローラ２０２－１は揮発性キャッシュメモリ２０８と不揮発性キャッシュメモリ２１０の両方にライトデータの書込み②を行い、キャッシュ管理テーブルを更新してホストコンピュータにライト完了を報告する。

【００１９】不揮発性キャッシュメモリ２１０に書込さ

れたデータは、予め定めたライトバック処理条件が成立した時、例えば一定期間、リード又はライトされなかった場合、ライトバック対象データとして抽出され、磁気ディスク装置に書込むライトバック③が行われる。このライトバック③に伴いライトバック済みのデータにつき不揮発性キャッシュ２１０を無効化して破棄させる。一方、ライトバック済みデータは揮発性キャッシュ２０８では有効であり、リード要求に対してはヒット状態にある。

【００２０】

【発明が解決しようとする課題】ところで、揮発性と不揮発性のキャッシュメモリを混在させた従来のディスクキャッシュ装置にあっては、揮発性キャッシュメモリのみの場合のネックであったライト処理時間が、リード処理でキャッシュヒットした場合と同じ時間に改善されている。

【００２１】しかしながら、不揮発性キャッシュメモリの容量が揮発性キャッシュメモリに比べて小さく、このためライトバックの回数が増加し、ライトバックのためにディスク側のコントローラと磁気ディスク装置の使用率が上昇し、システムのスループットが低下する問題がある。すなわち、不揮発性キャッシュメモリに格納できるデータ量が大きくとれないため、新規のライトデータを格納するには古いデータを磁気ディスク装置にライトバックして不揮発性キャッシュメモリを開放する必要が多く発生する。このため、ライトバック処理によりホストコンピュータからの新たな処理が実行出来ないことがある。

【００２２】また不揮発性キャッシュメモリ自体に何らかの障害が発生した場合には、ライトバックができなくなり、ディスクキャッシュの機能が失われてしまう問題があった。本発明は、このような従来の問題点に鑑みてなされたもので、磁気ディスクサブシステムの運用形態に応じた適合したキャッシュ機能を最大限に発揮できるディスクキャッシュ装置を提供することを目的とする。

【００２３】一方、近年にあっては、計算機システムの外部記憶装置として、記録の不揮発性、大容量性、データ転送の高速性等の特長を持つ磁気ディスク装置、光ディスク装置等のディスク装置が広く用いられている。ディスク装置に対する要求は、高速データ転送、信頼性重視、大容量性、低価格である。これらの要求を満たすものとして、ディスクアレイ装置が注目されてきている。ディスクアレイ装置とは、小型ディスク装置を数台から数十台並べ、複数のディスク装置に分散してデータを記録して、並列的にアクセスする装置である。

【００２４】ディスクアレイ装置で並列的に複数のディスク装置にデータ転送を行えば、一台のディスク装置の場合と比べて、ディスクの台数倍の高速データ転送が可能になる。また、データに加えて、パリティデータなどの冗長な情報を付け加えて記録しておくことで、ディスク装置の故障等を原因とするデータエラーの検出と訂正が可能となり、ディスク装置の内容を二重化して記録する方法と同程度の信頼性重視を、二重化より低価格で実現することができる。

【００２５】従来、カルフォルニア大学バークレイ校のデビット・Ａ・パターソン（DavidA. Patterson）らは、高速に大量のデータを多くのディスクにアクセスし、ディスク故障時におけるデータの冗長性を実現するディスクアレイ装置について、レベル１からレベル５までに分類付けを行って評価した論文を発表している（ACMSIGMOD Conferance, Chicago, Illinois, June 1-3, 1988 P109-P116）。

【００２６】このデビット・Ａ・パターソンらが提案したディスクアレイ装置を分類するレベル１～５は、ＲＡＩＤ（Redundant Arrays of Inexpensive Disks）１～５と略称される。ＲＡＩＤ１～５を簡単に説明すると次のようになる。ＲＡＩＤ０は、データの冗長性をもたないディスクアレイ装置であり、デビット・Ａ・パターソンらの分類には含まれていないが、これを仮にＲＡＩＤ０と呼ぶ。

【００２７】ＲＡＩＤ１は、２台のディスク装置を１組として同一データを書込むミラーディスク装置であり、ディスク装置の利用効率が低いが冗長性をもっており、簡単な制御で実現できるため、広く普及している。ＲＡＩＤ２は、データをビットやバイト単位でストライピング（分割）し、それぞれのディスク装置に並列に読み書きを行う。ストライピングしたデータは全てのディスク装置で物理的に同じセクタに記録する。

【００２８】データ用ディスク装置の他にハミングコードを記録するためのディスク装置を持ち、ハミングコードから故障したディスク装置を特定して、データを復元する。しかし、実用化されていてない。ＲＡＩＤ３は、データをビット又はバイト単位にストライピングしてパリティを計算し、ディスク装置に対しデータおよびパリティを並列的に書込む。ＲＡＩＤ３は、大量のデータを連続して扱う場合には有効であるが、少量のデータをランダムにアクセスするトランザクション処理のような場合には、データ転送の高速性が生かせず、効率が低下する。

【００２９】ＲＡＩＤ４は、１つのデータをセクタ単位にストライピングして同じディスク装置に書込む。パリティは固定的に決めたディスク装置に格納している。データ書込みは、書込み前のデータとパリティを読み出してから新パリティを計算して書き込むため、１度の書込みについて、合計４回のアクセスが必要になる。また書込みの際に必ずパリティ用のディスク装置へのアクセスが起きるため、複数のディスク装置の書込みを同時に実行できない。このようにＲＡＩＤ４の定義は行われているが、メリットが少ないため現在のところ実用化の動きは少ない。

【００３０】ＲＡＩＤ５は、パリティ用のディスク装置を固定しないことで、並列的なリード、ライトを可能にしている。即ち、セクタごとにパリティの置かれるディスク装置が異なっている。パリティディスクが重複しなければ異なるディスク装置にセクタデータを並列的に書込むことができる。このようにＲＡＩＤ５は非同期に複数のディスク装置にアクセスしてリード又はライトを実行できるため、少量データをランダムにアクセスするトランザクション処理に向いている。

【００３１】このようなＲＡＩＤ型で分類されるディスクアレイ装置は、システムの運用形態に応じてＲＡＩＤの動作形態が選択されているが、最近は、１つのディスクアレイ装置で、ＲＡＩＤ０，１，３および５に動作モードを異なるディスク装置に割当ててホストコンピュータが論理デバイスとして選択指定できるようにしたものが出されている。

【００３２】このようなＲＡＩＤ型のディスクアレイ装置においても、ディスクキャッシュ装置を使用することで処理性能をより一層向上させることができるが、ＲＡＩＤの動作形態に適合したディスクキャッシュの機能を最適化する必要がある。従って、本発明の他の目的は、ＲＡＩＤ型で動作形態が分類されたディスクアレイ装置を対象にディスクキッシュ機能の最適化を図ったディスクキャッシュ装置を提供するることを目的とする。

【００３３】

【課題を解決するための手段】図１は本発明の原理説明図である。

［第１発明］まず第１発明のディスクキャッシュ装置は、図１（Ａ）に示すように、上位装置１０からのアクセス要求に対しディスク装置３０のデータを読み書きする一対のディスク制御手段１２と、ディスク制御手段１２の各々に設けられ、ディスク装置３０に格納されたデータの一部を記憶するキャッシュ記憶手段４８と、一対のディスク制御手段１２の各々に設けられたキャッシュ制御手段４６と、一対のディスク制御手段１２の間を結合してアクセス要求などの情報およびデータの送受を行う共用バス手段１８とを備える。

【００３４】ここでキャッシュ制御手段４６には、信頼性重視モードキャッシュ制御手段７２、性能重視モードキャッシュ制御手段７４、又は平均モードキャッシュ制御手段７６が設けられ、動作モード設定手段７０により装置の立上り時に設定された動作モードに従ったキャッシュ制御が行われる。

I)　信頼性重視モード

動作モード設定手段７０で信頼性重視モードを設定した場合、一対のディスク制御手段１２に設けた信頼性重視モードキャッシュ制御手段７２の各々は、一対のディスク制御手段１２に設けたキャッシュ記憶手段４８の各々に、同一のデータを記憶する。

【００３５】具体的には、上位装置１０からのリード要求に対し自己のキャッシュ記憶手段４８に該当データがなかった場合（ミスヒット）は、ディスク装置３０から該当データを読出してステージングして上位装置１０にデータを転送し、同時に共通バス手段１８を介して他のディスク制御手段１２に転送して他方のキャッシュ記憶手段４８－２に同じデータをステージングさせる。

【００３６】また上位装置１０からのライト要求に対しては、自己のキャッシュ記憶手段４８にデータを書込むと共に、共通バス手段１８を介して他のディスク制御手段１２に転送して他方のキャッシュ記憶手段（４８）にも同じデータを書込ませる。更にライト要求の処理終了後のライトバック処理として、ライトバック条件の成立時に、自己のキャッシュ記憶手段４８からディスク装置３０への書込みが済んでいないデータ（以下「ダーティデータ」という）を抽出してディスク装置３０に書き戻すと共に、共用バス手段１８を介して書戻し済みデータのキャッシュ無効化を他方のディスク制御手段１２に指示する。

II)　性能重視モード

動作モード設定手段７０で性能重視モードを設定した場合は、一対のディスク制御手段１２に設けた性能重視モードキャッシュ制御手段７４の各々は、一対のディスク制御手段１２に設けたキャッシュ記憶手段４８の各々に、自己にアクセス要求のあったデータのみを個別に記憶する。

【００３７】具体的には、上位装置１０からのリード要求に対し自己のキャッシュ記憶手段４８に該当データがなかった場合は、他方のディスク制御手段１２に対し共用バス手段１８を介して処理を依頼し、他のディスク制御手段１２からのデータなしの応答を受けた際にディスク装置３０から該当データを読出して自己のキャッシュ記憶手段４８にステージングして上位装置１０にデータを転送する。

【００３８】また上位装置１０からのライト要求に対しては、自己のキャッシュ記憶手段４８にのみデータを書込む。ライトバック処理は、書戻し条件が成立した際に、自己のキャッシュ記憶手段４８からディスク装置３０へ書込みが済んでいないダーティデータを抽出してディスク装置３０に書き戻す。

III)　平均モード

動作モード設定手段７０で平均モードを設定した場合、一対のディスク制御手段１２に設けた平均モードキャッシュ制御手段７６の各々は、一対のディスク制御手段１２の各々は、上位装置１０からのリード要求に伴うディスク装置３０からのステージングは自己のキャッシュ記憶手段４８に対してのみ行い、上位装置１０からのライト要求に伴うデータの書込みは、一対のディスク制御手段１２のキャッシュ記憶手段４８の各々に行って同一データを記憶する混在型とする。

【００３９】具体的には、上位装置１０からのリード要

求に対し自己のキャッシュ記憶手段４８に該当データがなかった場合は、他方のディスク制御手段１２に対し共用バス手段１８を介して処理を依頼し、依頼後にディスク装置３０から該当データを読出して自己のキャッシュ記憶手段４８にステージングする。ここで他方のディスク制御手段１２への処理依頼に対しデータなしの応答を受けていた際には、ディスク装置３０から該当データを読出して自己のキャッシュ記憶手段４８にステージングした後に上位装置１０に転送する。

【００４０】また上位装置１０からのライト要求に対しては、自己のキャッシュ記憶手段４８にデータを書込むと共に、共通バス手段１８を介して他のディスク制御手段１２に転送して他方のキャッシュ記憶手段４８にも同じデータを書込ませる。この点は信頼性モードと同じである。更にライトバック処理は、書戻し条件が成立した際に、自己のキャッシュ記憶手段４８からディスク装置３０への書込みが済んでいないダーティデータを抽出してディスク装置３０に書き戻すと共に、共用バス手段１８を介して書戻しデータのキャッシュ無効化を他方のディスク制御手段１２に指示する。これは性能重視モードと同じである。

【００４１】IV) メモリ構成

更に、信頼性重視モード、性能重視モード及び平均モードのいずれについても、一対のディスク制御手段１２に設けたキャッシュ記憶手段４８の各々を、不揮発性メモリとする。但し、信頼性重視モードについては、両方とも揮発性メモリでもよい。またディスク装置３０をディスクアレイ装置としてもよい。

［第２発明］図１（Ｂ）はディスクアレイ装置のＲＡＩＤ形態に適合した第２発明のディスクキャッシュ装置の原理説明図である。

【００４２】このディスクキャッシュ装置は、複数のディスク装置３０を設けたディスクアレイ手段２８を備える。ディスクキャッシュ機構としては、ディスクアレイ手段２８に格納されたデータの一部を記憶して上位装置１０からのアクセス要求に対しデータを読み書きするキャッシュ記憶手段４８を持ち、キャッシュ管理手段５０のハッシュテーブル手段６０およびＬＲＵテーブル手段６２に基づいてキャッシュ記憶手段４８の記憶状態を管理する。

【００４３】またリードキュッシュ制御手段４６－１とライトキャッシュ制御手段４６－２が設けられる。リードキャッシュ制御手段４６－１は、上位装置１０からリード要求を受けた際に、キャッシュ管理手段５０のハッシュテーブル手段６０を参照し、要求データが存在する場合は、キャッシュ記憶手段４８から読出して上位装置１０に転送する。要求データが存在しない場合は、ディスクアレイ手段２８からステージングした後に上位装置１０に転送する。

【００４４】ライトキャッシュ制御手段４６－２は、上位装置１０からライト要求を受けた際に、キャッシュ管理手段５０のハッシュテーブル手段６０に書込を示す情報を登録すると共に、キャッシュ記憶手段４８にデータを書込み、書込み済みのキャッシュブロックをＬＲＵテーブル手段６２の先頭にリンクする。またライトバック制御手段が設けられ、予め定めたライトバック条件が成立した際に、キャッシュ記憶手段４８からディスクアレイ手段２８への記憶が済んでいないデータを抽出して書戻す。

【００４５】更にディスクアレイ制御手段５０が設けられ、リードキャッシュ制御手段４６－１および前記ライトキャッシュ制御手段４６－２によるディスクアレイ手段２８のアクセス時に、予め設定されたディスクアレイのＲＡＩＤ動作モードに従って１又は複数のディスク装置３０のアクセスを制御する。

I) ＲＡＩＤ動作モード

ディスクアレイ手段２８は、並列的に配置された入出力ポートの各々にディスク装置３０を接続して１ランクを構成し、この１ランク構成のディスク装置群を複数ランク設けている。ディスクアレイ制御手段５２は、ＲＡＩＤ０，ＲＡＩＤ１、ＲＡＩＤ３又はＲＡＩＤ５に対応した第１～第４動作モードを設定する。

【００４６】ＲＡＩＤ０に対応した第１動作モードの設定時は、特定のランクの中のディスク装置３０を上位装置１０で扱う論理デバイスに１対１に割当て、個々のディスク装置３０に読み書きを行う。ＲＡＩＤ１に対応した第２動作モードの設定時は、特定のランクの中の２台のディスク装置３０を１組として上位装置１０で扱う論理デバイスに１対１に割当て、２つのディスク装置３０に同一データを書込むと共にいずれか一方のディスク装置３０からデータを読出すミラーディスクとする。

【００４７】ＲＡＩＤ３に対応した第３動作モードの設定時は、特定のランクを構成するｎ台のディスク装置３０をデータ用とパリティ用とに固定的に割当て、上位装置１０からの書込要求時に、書込データをビット単位又はバイト単位に（ｎ－１）分割すると共に分割単位ごとにパリティデータを計算し、（ｎ－１）分割したデータおよびパリティデータをｎ台のディスク装置３０に並列的に書込む。

【００４８】ＲＡＩＤ５に対応した第４動作モードの設定時は、特定のランクを構成するｎ台のディスク装置３０を１組としてアクセスごとにパリティ位置が変化するように割当て、上位装置１０からの書込要求時に、書込データを少なくともセクタ単位に分割すると共に、書込データの書込先ディスク装置３０およびパリティ用ディクス装置から旧データおよび旧パリティデータを読出して新パリティを計算し、前記書込データ及び新パリティデータを前記書込先ディスク装置３０およびパリティ用ディクス装置に並列的に書込む。

II) キャッシュメモリの空スペースの確保

キャッシュ管理手段５０は、キャッシュ記憶手段４８の空き状態を管理する空スペース管理テーブル手段６０を備える。ライトキャッシュ制御手段４６－１は上位装置１０に基づくキャッシュ書込処理の完了報告時に空スペース管理テーテブル手段６４を参照し、空スペースが予め定めた閾値以下の場合、ディスク書込済みのデータは無効化してキャッシュ記憶手段４８から削除し、且つディスク書込みが済んでいないダーティデータはライトバック手段４５による処理動作を準備する。これにより常に一定量の空スペースがキャッシュ記憶手段４８に確保され、ライト動作が迅速にできる。

【００４９】空スペース管理テーブル手段６４には、キャッシュ記憶手段４８上の連続するキャッシュブロックの空ブロック数をエントリ（インデックス）として先頭ブロックの位置情報を格納する。

III）ステージング動作]

リードキャッシュ制御手段４６－１は、上位装置１０からのリード要求データがキャッシュ記憶手段４８の格納単位であるキャッシュブロック上に存在しない時、キャッシュブロック上に存在する有効データの状態に応じてディスクアレイ手段２８から新たなデータのステージングを行って連続する一つのデータ領域を生成する。

【００５０】ステージング終了後に、リード要求データに続く有効データをディスクアレイ手段２８からキャッシュブロック上にプリフェッチする。このプリフェッチするデータ量は、ＲＡＩＤ１に対応する第３動作モードの設定時は、予め設定した量の有効データとする。ＲＡＩＤ３に対応する第３動作モードの設定時又はＲＡＩＤ５に対応する第４動作モードの設定時には、キャッシュブロックの最後までの有効データとする。

【００５１】ＲＡＩＤ５に対応する第４動作モードの設定時には、リード要求データを含むパリティグループの最後までの有効データとしてもよい。これはキャッシュブロックサイズとパリティグループのブロックサイズが一致する場合である。ＲＡＩＤ５に対応する第４動作モードの設定時に、リードキャッシュ制御手段４６－１は、オプション指定によりキャッシュ記憶手段４８にパリティデータもステージングし、この場合、ライトバック手段４５は、キャッシュ記憶手段４８のパリティデータを用いて新たなパリティを計算する。これによりライトバックの際の旧データおよび旧パリティのディスク読出を不要にできる。

IV）シーケンシャルデータのＬＲＵ管理

リードキャッシュ制御手段４６－１は、上位装置１０から大量シーケンシャルデータのリード要求に対しステージングを行った場合、キャッシュ記憶手段４８のＬＲＵテーブル手段６２を参照して末尾から所定量のキャッシュブロックとなる記憶データの一部を無効化すると共に、大量シーケンシャルデータをＬＲＵテーブル手段６２の先頭に登録せずに終端側に登録する。これによって比較的短期間で大量シーケンシャルデータをキャッシュ記憶手段から追い出すことができる。

V）ライトバック処理]

ＲＡＩＤ５に相当する第４動作モードの設定時に、ライトバック手段４５は、キャッシュブロックの中のライトバックの対象となった有効データをパリティグループ単位に抽出してパリティデータを算出した後に、抽出した有効データおよびパリティデータをディスクアレイ手段２８の複数のディスク装置３０に並列的に書込む。即ち、ＲＡＩＤ５であってもＲＡＩＤ３と同様に動作させてライトバックを効率良く行う。

【００５２】またライトバック手段４５は、キャッシュブロックの中のライトバックの対象となった有効データのパリティグループに対する不足数が所定数以下の場合、例えば１ブロック以下の場合、不足データを対応するディスク装置３０から読出した後にパリティグループ分の有効データおよび計算したパリティデータをディスクアレイ手段２８の複数のディスク装置３０に並列的に書込むことでＲＡＩＤ３的な動作を可能とする。

【００５３】ライトバック手段４５は、所定のライトバック条件が成立した際に、ＬＲＵテーブル手段６２内の終端から所定数のキャッシュブロックをライトバック対象に指定すると共に、この範囲に含まれるキャッシュブロックをデータが連続するように並び替え、ＲＡＩＤ３的なライトバックを可能とする。

VI）キャッシュブロックサイズの自動設定]

更に本発明は、装置の使用中に上位装置１０からのアクセス要求で使用されるデータブロックサイズを統計情報として収集し、この統計情報の平均値を次の装置立ち上げ時のキャッシュ記憶手段４８のキャッシュブロックサイズとして自動設定する統計処理手段を設ける。

【００５４】

【作用】まず図１（Ａ）の第１発明の作用を説明する。第１発明のディスクキャッシュ装置で信頼性重視モードを設定した場合には、、両系統のキャッシュメモリ内のデータを常に同一とするようにキャッシュ動作が行われる。即ち二重化されたキャッシュメモリを独立に使用するのではなく、片側データをバックアップ用と考えている。

【００５５】この場合の長所は、データが常に二重化されており片側データの喪失時データが保障され高い信頼性が得られる。また短所としては、キャッシュメモリの半分がバックアップ用であり、使用効率が低くなる。また第１発明のディスクキャッシュ装置で性能重視モードを設定した場合には、各キャッシュメモリのデータは独立であり、信頼性モードの最大２倍の容量のキャッシュメモリとして使用できる。

【００５６】この場合の長所は、両系統のキャッシュメモリが独立に動作し、また他方からの参照更新が可能であるためキャッシュ容量が増大する。短所としては、メ

モリ障害，電源障害でデータ消失、データ化け等の障害が発生する確立が高くなる。また他系統にキャッシュ処理を依頼したヒットとなった時は、リモート側キャッシュに対してのリードとなるので性能が低下する。

【００５７】更に第１発明のディスクキャッシュ装置で平均的モードを設定した場合は、基本的には信頼性重視モードと性能重視モードの混在型となる。すなわち、リード要求のミスヒット時のステージングは自己のキャッシュメモリに対してのみ行い、またデータのライトは両方系統のキャッシュメモリに対して行う。これによりライト処理の少ない環境ではキャッシュ容量は性能重視モードと同様に２倍となり、またライト処理が多く発行される環境では信頼性重視モードと同様に障害に対して強い形になる。この場合の長所と短所は、使用環境、即ちリードとライトの発行比率に応じていずれかの信頼性重視モードと性能重視モードの長所と短所が強く現れることになる。

【００５８】次に図１（Ｂ）のディスクアレイを対象としたディスクキャッシュ装置の作用を説明する。まず第２発明にあっては、キャッシュメモリの管理に使用する空スペース管理テーブルを備える。すなわちキャッシュメモリ上にないデータに対してのステージング又はライトが指示された場合、コントローラはキャッシュメモリ上の未使用の空間から新たなキャッシュブロックを割当て、ディスク装置からのステージングによるデータ、あるいはホストコンピュータからのライトデータを受領する。

【００５９】この場合、キャッシュメモリ上に一定量の空き空間が残るように空スペース管理テーブルを用いて管理される。すなわち残り容量が規定値以下になるような状況が発生すると、キャッシュメモリ上の有効データの追い出しがスケジュールされる。即ち、ディスク書込が済んだデータはハッシュエントリで指定されるＬＲＵテーブルの登録を抹消する無効化を行う。またディスク書込みが済んでいないダーティデータは、ディスクへの書き戻しがスケジュールされる。これにより新たな上位装置からのリード又はライト要求に対し常にキャッシュメモリ上に空き空間が用意され、直ちに処理が開始できる。

【００６０】次に第２発明にあっては、リード要求に対しキャッシュメモリがミスヒットとなった時のディスク装置からのステージングを、キャッシュメモリを使用したことによる効果が最大に発揮できるように行う。これは要求されたホストブロック分のデータをディスク装置からキャッシュブロック上にステージングした後に、後続するホストブロックに対応するデータをディスク装置からプリフェッチすることで実現される。

【００６１】このようにステージングが済んだホストブロックに引き続くブロックに対して先読み動作を行うことで、ホストコンピュータからの逐次リード要求に対してキャッシュ効果を最大限にする処理ができる。この場合のプリフェッチ量は、ＲＡＩＤ０，１，３，５で異なる。まずＲＡＩＤ３は１つのホストブロックが所定バイト数単位に各ディスク装置に分散して格納されており、ステージングが高速に行われること及び大容量データが要求される場合が多いことを考慮し、ステージングされた有効ブロックを含むキャッシュブロックの最後までのホストブロックに対応するデータをプリフェッチする。

【００６２】ＲＡＩＤ５は、基本的にはリード転送速度はディスク装置の速度以上は出ず、、ステージング速度がＲＡＩＤ３より遅いこと、また小容量データのリード処理が中心になることから、ステージングが済んだブロックに続く次のパリティグループ（パリティ用ディスク装置を同一とするグループ）までとする。ここで、キャッシュブロックサイズとパリティグループのブロックサイズが同一となるようにストライピングを行えば、キャッシュブロックの最後迄がプリフェッチ量となり、ＲＡＩＤ３的な高速ステージングが可能となる。

【００６３】更にプリフェッチ後のキャッシュブロック上で、連続する有効ブロックに空きがあって分散している場合には、この空ブロックに内部的なステージングを実行し、キャッシュブロック上の有効データを連続ブロック領域とし、データ管理を容易にする。また第２発明のディスクキャッシュ装置は、大量シーケンシャルデータのステージングに対し、その追い出しを可能ながぎり早めるようにキャッシュメモリを管理する。

【００６４】一般にディスクキャッシュ装置は同一データを頻繁にリードすると効果が出る。ホストコンピュータがキャッシュメモリ上に存在しない大量データのリード要求を行ったとき、ディスク装置からデータをステージングしホストに転送する。このデータは一度しか参照されないにも関わらずそのままキャッシュメモリ上に残しておくと、キャッシュスペースが有効に利用されないことになる。しかし直ちに無効化することも出来ない。

【００６５】そこで、ホストコンピュータが一定量以上のデータ、例えばキャッシュブロック１００分相当等を要求した時、ステージング後、本来はＬＲＵテーブルの先頭位置に登録すべきところを、第２発明は、ＬＲＵテーブルの終端の近傍に登録する。これにより比較的短い時間で、大量シーケンシャルデータは擬似的に最も使用されていない古いデータとなり、キャッシュ無効化により追い出すことができる。

【００６６】また第２発明は、オプション指定によりＲＡＩＤ５でディスクアレイからパリティデータを含めてキャッシュメモリにステージングすることができる。このようなパリティデータを含めたステージングにより、ライトバック処理でディスクアレイからの旧データおよび旧パリティデータのリードが不要となり、ＲＡＩＤ５のもつライトペナルティを軽減させることができる。

【００６７】具体的には、パリティのステージングを行

わない場合、ディスク装置の２．５回転分の時間を必要とするが、パリティをステージングしていた場合には、１．５回転分の時間で済む。但し、キャッシュメモリの容量はパリティをステージングした分だけ減少する。さらに第２発明のディスクキャッシュ装置は、ＲＡＩＤ５の動作状態で、ライト処理後に行なうディスクアレイに対するライトバック処理を最適化する。

【００６８】ＲＡＩＤ５では、パリティグループを構成する全てのディスク装置に跨がって連続するデータがキャッシュブロック上に存在する場合は、ＲＡＩＤ３的に連続データからパリティを計算し、データおよびパリティの同時転送で並列的に書込むことができる。一方、パリティグループを構成する一部のディスク装置のデータしかキャッシュブロック上に存在しない場合は、旧パリティと旧データをリードしてパリティを計算した後にデータとパリティを並列的に書込む通常のリードモディファイライトを行う。

【００６９】この場合、キャッシュブロック上で連続するデータの空きが所定値以下、例えば１ディスク装置分であった場合には、空部分に有効データをステージングして連続データとした後に、ＲＡＩＤ３的な同時更新のライトバックを行う。更に、キャッシュメモリの空きスペースが規定値以下となってライトバック処理がスケジュールされた場合、ＬＲＵテーブルの末尾から一定範囲のキャッシュブロックをライトバック対象に指定する。このライトバック対象として指定されたキャッシュブロックに対し、ホストブロックが連続するように並び替えを行い、ＲＡＩＤ３的な高速のライトバック処理を可能とする。

【００７０】更に第２発明のディスクキャッシュ装置は、ユーザの運用形態に応じて自動的にキャッシュブロック・サイズを設定することができる。一般にユーザシステムにおいてユーザが必ず設定を必要とるパラメータは、ホストブロックサイズ（ロジカルブロックサイズ）である。キャッシュブロックサイズは、ホストブロックサイズが複数集まったサイズとして設定される。第２発明は、最初のシステム立ち上げ時は、ディフォルトのキャッシュブロックサイズで初期動作するが、次の立ち上げ時には、ユーザの使用環境に応じて自動的にキャッシュブロックサイズを変更する。

【００７１】即ち、ユーザ使用中にキャッシュシステムをアクセスしたホストブロックサイズの統計を取り、平均的なホストブロックサイズを求めている。そして次回の電源投入に伴う立ち上げ時に、前回の運用で求めた新規なキャッシュブロックサイズでキャッシュシステムを動作させ、キャッシュブロック・サイズをユーザの判断を必要とすることなく最適化する。

【００７２】

【実施例】

＜目次＞

１．システムのハードウェア構成

図２は本発明のディスクキャッシュ装置が適用される磁気ディスク装置を用いた入出力サブシステムのハードウェア構成を示す。

【００７３】上位装置としてのホストコンピュータ１０には少なくとも２つのチャネル装置１４－１，１４－２が設けられ、チャネル１４－１，１４－２に対し、チャネルインタフェース１６を介して２台のコントローラ１２－１，１２－２を接続している。この実施例で、チャネルインタフェース１６としてはＳＣＳＩを使用している。勿論、ＭＢＣインタフェース（ブロック・マルチプレクサ・チャネルインタフェース）を使用してもよい。

【００７４】コントローラ１２－１，１２－２はディスク制御手段としての機能を有し、キャッシュ機構を内蔵している。コントローラ１２－１，１２－２のデバイス側は共用バス１８－１，１８－２に接続される。共用バス１８－１，１８－２はブリッジ回路部２０で結合され、コントローラ１２－１，１２－２間でメッセージ、コマンド等の情報およびデータのやり取りを行うことができる。更に共用バス１８－１，１８－２はサブコントローラ２２－１，２２－２が接続され、コントローラ１２－１，１２－２の処理機能を分散させて負荷の低減を図っている。

【００７５】共用バス１８－１，１８－２にはアダプタ２４－１～２４－６，２６－１～２６－６のそれぞれを介して、ディスクアレイ２８に設けている複数のディスク装置３０－００～３０－３５が接続される。この実施例でディスクアレイ２８はコントローラ１２－１，１２－２より並列アクセスを受ける６つのポートＰ０～Ｐ５で並列ディスク群を構成し、この並列ディスク群をランクＲ０～Ｒ３で示す４ランク分設けている。

【００７６】具体的には、ランクＲ０はポートＰ０～Ｐ５に対応した６台のディスク装置３０－００～３０－０５で構成され、ランクＲ１はポートＰ０～Ｐ５に対応したディスク装置３０－１０～３０－１５で構成され、ラ

ンクＲ２はポートＰ０～Ｐ５に対応したディスク装置３０－２０～３０－２５で構成され、更にランクＲ３はポートＰ０～Ｐ５に対応したディスク装置３０－３０～３０－３５で構成される。

【００７７】このようなディスクアレイを構成するディスク装置の位置はランク番号Ｒとポート番号Ｐアドレスで定義する。例えば磁気ディスク装置３０－００は（Ｒ０，Ｐ０）で表わすことができる。図３は図２のコントローラ１２－１側のハードウェア構成を示す。コントローラ１２－１内にはＣＰＵ３２が設けられ、ＣＰＵ３２の内部バス４４にＲＯＭ３４、ＤＲＡＭ３６、ＳＣＳＩ回路部４０とのやり取りを行う上位インタフェース部３８、共用バス１８－１とのやり取りを行うバスインタフェース部４２が設けられる。

【００７８】更にディスクキャッシュ機構を実現するためキャッシュ制御部４６とキャッシュメモリ４８が設けられる。なお、キャッシュ制御部４６を独立に設けて、ＣＰＵ３２の機能として実現してもよい。またキャッシュ制御部４６においてキャッシュメモリ４８の管理に使用されるキャッシュ管理テーブル、具体的にはキャッシュエントリテーブル，ＬＲＵテーブルおよび空スペース管理テーブルはＤＲＡＭ３６に展開されている。

２．第１発明の全体機能

図４は図２のハードウェア構成を対象とした第１発明の処理機能を示す。尚、説明を簡単にするため、ディスクアレイ側については１台のディスク装置３０を代表して示し、共用バスについても１本の共用バス１８で示している。またキャッシュメモリについてはコントローラから取出して示している。

【００７９】コントローラ１２－１，１２－２には動作モード設定部７０－１，７０－２、信頼性重視モードキャッシュ制御部７２－１，７２－２、性能重視モードキャッシュ制御部７４－１，７４－２、および平均モードキャッシュ制御部７６－１，７６－２が設けられている。キャッシュメモリとしては不揮発性キャッシュメモリ４８－１，４８－２が使用される。

【００８０】ここでコントローラ１２－１側を例にとると、動作モード設定部７０－１は予め定めたホストコンピュータ１９からのコマンド指示あるいはコントローラ１２－１に予め設けられたピン設定などにより、信頼性重視モード，性能重視モードまたは平均モードのいずれか１つを設定する。この動作モード設定部７０－１による設定に基づき、信頼性重視モードキャッシュ制御部７２－１，７２－２、性能重視モードキャッシュ制御部７４－１，７４－２、または平均モードキャッシュ制御部７６－１，７６－２のいずれか１つの動作を有効とする。

【００８１】ここでホストコンピュータ１０に対し２台のコントローラ１２－１，１２－２が設けられていることから、ホストコンピュータ１０からのアクセス要求を受けたコントローラ側をローカル側といい、このときアクセス要求を受けていない側をリモート側という。これはコントローラ１２－１，１２－２において相対的なものであり、コントローラ１２－１に対しホストコンピュータ１０からアクセス要求が行われたときはコントローラ１２－１がローカル側、コントローラ１２－２がリモート側となる。

【００８２】逆にコントローラ１２－２に対しホストコンピュータ１０からアクセス要求が行われたときはコントローラ１２－２がローカル側、コントローラ１２－１がリモート側となる。

３．信頼性重視モードのディスクキャッシュ制御

信頼性重視モードにあっては、コントローラ１２－１，１２－２に設けたキャッシュメモリは図４に示したように両方とも不揮発性キャッシュメモリ４８－１，４８－２とする。

【００８３】尚、信頼性重視モードにあっては、キャッシュメモリを両方とも揮発性としてもよい。この場合、後の説明で明らかにする性能重視モードおよび平均モードにあっては、両方とも不揮発性キャッシュメモリとする必要があるため、電源バックアップの有無により不揮発性となるか揮発性と不揮発性が切替可能なメモリを使用する必要がある。

【００８４】信頼性重視モードにあっては、コントローラ１２－１，１２－２の各不揮発性キャッシュメモリ４８－１，４８－２のデータが同一であるようにキャッシュ動作が行われる。即ち、２系統のキャッシュメモリを独立に使用するのではなく、同一データを格納することで二重化し、片側のデータ送出時にあっても、キャッシュデータを保証して信頼性を高める。

【００８５】リード要求に対するミスヒット時のステージングはディスク装置３０から両方の不揮発性キャッシュメモリ４８－１，４８－２に対し行う。またライト要求に対するライトデータの書込みも両方の不揮発性４８－１，４８－２に対し行う。更に、ライト処理終了後のディスク装置３０に対するライトバックについては、ライト要求を受けたローカル側のキャッシュメモリから行う。

【００８６】図５は信頼性重視モードにおけるキャッシュ動作を示したフローチャートである。まずステップＳ１にあっては、ホストコンピュータ１０からのアクセス要求の有無を監視しており、アクセス要求があるとステップＳ２に進み、リード要求かライト要求かのアクセスモードを判別する。リード要求であれば、ステップＳ３で自己の不揮発性キャッシュメモリ即ちローカル側キャッシュメモリに要求データが存在するか否かのヒット判定を行う。

【００８７】キャッシュヒットが得られれば、ステップＳ４で要求データを読み出してホストコンピュータ１０にデータ転送する。キャッシュミスヒットの場合にはデ

ィスク装置３０からのステージングを行うことになるが、ステージングに先立ち、リモート側のコントローラ１２－１にステージングを通知する。このステップＳ５における処理は図７に示される。即ち、ホストコンピュータ１０からのリード要求を受けたコントローラ１２－１が不揮発性キャッシュメモリ４８－１のミスヒットを判定すると、ディスク装置３０からのステージングに先立ってローカル側のコントローラ１２－２に対し共用バス１８を経由してメッセージを送る。

【００８８】このメッセージはキャッシュミスヒットとなったため、ステージングによりディスク装置３０からのデータをリモート側の不揮発性キャッシュメモリ４８－２にも転送することを通知する。このローカル側のコントローラ１２－１からのメッセージを受けてリモート側のコントローラ１２－２は自己のキャッシュ管理テーブルを更新し、不揮発性キャッシュメモリ４８－２内にステージング用の空間を割り当て、スタンバイ状態となったことをローカル側のコントローラ１２－１に送り返す。

【００８９】再び図５を参照するに、ステップＳ５でリモート側へのステージングの通知が済むと、リモート側からのレディ応答を待って、ステップＳ６で要求データをディスク装置３０から読み出し、ローカル側の不揮発性キャッシュメモリ４８－１に格納すると共に、共用バス１８を介してリモート側の不揮発性キャッシュメモリ４８－２に転送し、それぞれ書き込むステージングを行う。

【００９０】ローカル側の不揮発性キャッシュメモリ４８－１に対するステージングが済むと、ステージングにより得られた要求データ④をホストコンピュータ１０に転送し、ステップＳ８でリード終了応答を行って一連のリード要求に対する処理を終了する。次にステップＳ２でライト要求を判別した場合には、ライト要求を受けたローカル側のコントローラ１２－１がリモート側のコントローラ１２－２へ共用バス１８を介してライトデータの転送を通知する。

【００９１】例えば図８に示すように、ホストコンピュータ１０よりライト要求①をコントローラ１２－１が受けると、共用バス１８を経由してリモート側のコントローラ１２－２に対しメッセージ②を送る。このメッセージ②はライト要求により不揮発性キャッシュメモリ４８－２の更新が必要であり、ライトデータを共用バス１８を経由してリモート側の不揮発性キャッシュメモリ４８－２に転送することを通知する。

【００９２】このメッセージ②を受けたリモート側のコントローラ１２－２は自己のキャッシュ管理テーブルを更新し、不揮発性キャッシュメモリ４８－２内に書込用の空間を割り当て、レディ応答をローカル側のコントローラ１２－１に返す。リモート側コントローラ１２－２からのレディ応答をステップＳ１０で判別すると、ローカル側のコントローラはホストコンピュータ１０からのライトデータ③を自己の不揮発性キャッシュメモリ４８－１、即ちローカル側のキャッシュメモリ４８－１に書き込み、同時にステップＳ２でリモート側のキャッシュメモリ４８－２に共用バス１８経由で転送して書き込ませる。

【００９３】この結果、ライトデータはローカル側およびリモート側のキャッシュメモリ４８－１，４８－２の両方に書き込まれる。ステップＳ１２でローカル側およびリモート側のキャッシュメモリ４８－２に対するライトデータの転送書込みが済むと、この時点でライトバックは行わず、ステップＳ１３でライト終了応答をホストコンピュータ１０に転送して、一連のライト要求に対する処理を終了する。

【００９４】図６は信頼性重視モードにおけるライトバックの処理動作を示したフローチャートである。図５のライト要求に対する処理に示したように、ローカル側およびリモート側のキャッシュメモリにライトデータが書き込まれた時点ではディスク装置に対する書込み所謂ライトスルーは行わず、予め定めたライトバック条件が成立したときにディスク装置に書き戻すライトバック処理を実行する。

【００９５】このライトバック処理は、まずステップＳ１でコントローラ１２－１がアイドル状態か否かチェックしており、アイドル状態になければステップＳ６で要求された事項の処理を行っているが、アイドル状態に入るとステップＳ２でライトバック条件の成立の有無をチェックする。ここでライトバック条件とはキャッシュ管理テーブルに設けているＬＲＵテーブルに基づきキャッシュデータの追い出し条件がスケジュールされている場合である。

【００９６】即ち、キャッシュ管理テーブルのＬＲＵテーブルには、最新に使用されたキャッシュブロックを先頭とし、最も使用されていない時間のたったキャッシュブロックを末尾とするキャッシュブロックの登録が行われており、例えば一定時間使われなかったテーブル末尾のキャッシュブロックがキャッシュ無効化されるが、この内ディスク装置への書込みが済んでいないデータ、所謂ダーティデータについてはライトバックがスケジュールされる。

【００９７】ステップＳ２でライトバック条件の成立を判定すると、ライトバック対象となっているＬＲＵテーブルの末尾から一定範囲のキャッシュブロックを読み出してディスク装置３０に書き込むライトバックを実行する。例えば図８に示したように、先行するライト処理でコントローラ１２－１がホストコンピュータ１０からライト要求①を受けて自己の不揮発性キャッシュメモリ４８－１にライトデータを書き込んでおり、その後のアイドル状態でライトバック条件が成立した場合には、④に示すようにディスク装置３０にデータを書き込むライト

バック処理を行う。

【００９８】再び図６を参照するに、ステップＳ３でディスク装置３０に対するライトバックが済むと、ステップＳ４でライトバックが済んだキャッシュブロックをキャッシュ管理テーブルのハッシュテーブルから削除するキャッシュ無効化を行う。更に、ローカル側のコントローラ１２－１はステップＳ５で共用バス１８を経由してリモート側のコントローラ１２－２にメッセージを送り、ライトバックが済んだデータのキャッシュ管理テーブルからの削除を指示し、リモート側についてもキャッシュ無効化を行う。

４．容量重視モードのキャッシュ制御

この性能重視モードは２台のコントローラ１２－１，１２－２に設けた各キャッシュメモリ４８－１，４８－２は不揮発性とし、各キャッシュメモリ４９８－１，４８－２のデータを独立とする。そのため、２つのキャッシュメモリ４８－１，４８－２に同一データを格納した信頼性重視モードに比べ、最大で２倍のキャッシュ容量とすることができる。換言すると、性能重視モードは容量重視モードと言うことができる。

【００９９】リード処理はリード要求を受けたローカル側でミスヒットとなった場合にはリモート側にキャッシュ処理を依頼し、リモート側でキャッシュヒットになれば、直接、要求データをホストコンピュータにデータ転送する。ディスク装置からのステージングはローカル側およびリモート側の両方でミスヒットとなった場合にディスク装置からローカル側に対してのみ行う。またライト処理については、ライト要求を受けたローカル側のキャッシュメモリに対してのみ行う。

【０１００】更にライトバックについても、ライトバック条件が成立したローカル側のキャッシュメモリからディスク装置に書き込むライトバックのみ行い、信頼性重視モードにおけるライトバック終了後のリモート側に対するキャッシュ無効化の指示は行わない。図９は性能重視モードにおけるキャッシュ動作を示したフローチャートである。ステップＳ１でホストコンピュータ１０からアクセス要求があると、ステップＳ２でアクセスモードを判別し、リード要求であればステップＳ３で自己のキャッシュメモリ４８－１、即ちローカルキャッシュメモリ４８－１のヒットの有無を判定する。

【０１０１】ローカルキャッシュメモリ４８－１のヒットが判別されると、ステップＳ４で要求データをホストコンピュータ１０にデータ転送する。ローカル側キャッシュメモリ４８－１でミスヒットとなった場合には、ステップＳ５で共用バス１８を経由してリモート側のコントローラ１２－２にキャッシュ処理を依頼する。このローカル側からのキャッシュ処理の依頼を受けたリモート側のコントローラ１２－２は、リモート側キャッシュメモリ４８－２でキャッシュヒットになれば要求データを直接ホストコンピュータ１０にデータ転送し、応答結果をローカル側のコントローラ１２－１に通知してくる。

【０１０２】一方、リモート側でもキャッシュミスヒットとなった場合には、ステップＳ６でリモート側からのヒット応答が得られないことからステップＳ７に進み、ディスク装置３０からローカル側のキャッシュメモリ４８－１に要求データを書き込むステージングを行う。ステージングが済むと、ステップＳ８でホストコンピュータ１０へのデータ転送を行い、これ以降ステージングされたデータはヒット可能状態となる。このような一連の処理を終了すると、リモート側のコンローラ１２－１はステップＳ９でホストコンピュータ１０にリード終了応答を返す。

【０１０３】図１１は図９のステップＳ５～Ｓ８に示したローカル側およびリモート側の両方の不揮発性キャッシュメモリ４８－１，４８－２がホストコンピュータ１０からのリード要求①に対しミスヒットとなったときの動作を示している。即ち、ホストコンピュータ１０からのリード要求①を受けたローカル側のコントローラ１２－１は自己の不揮発性キャッシュメモリ４８－１のミスヒット②を判定すると、共用バス１８を経由してリモート側のコントローラ１２－２にメッセージを送り、リモート側の不揮発性キャッシュメモリ４８－２に要求データが存在するか否かの調査を依頼する。

【０１０４】リモート側のコントローラ１２－２は自己のキャッシュ管理テーブルのハッシュテーブルを参照し、もし要求データが存在すればヒット応答と要求データが存在するアドレスをローカル側のコントローラ１２－１に通知する。これに基づき、ローカル側のコントローラ１２－１はリモート側の不揮発性キャッシュメモリ４８－２から要求データ⑤を読み出してホストコンピュータ１０に転送する。

【０１０５】尚、図１１にあっては、リモート側の不揮発性キャッシュメモリ４８－２から読み出した要求データ⑤をコントローラ１２－２から直接ホストコンピュータ１０にデータ転送しているが、共用バス１８からローカル側のコントローラ１２－１を経由してホストコンピュータ１０にデータ転送をすることもできる。しかし、リモート側のコントローラ１２－２から直接転送した方が、新たなホストコンピュータ１０とコントローラ１２－２との間のチャネル結合を考慮しても高速に処理できる。

【０１０６】再び図９を参照するに、ステップＳ２でライト要求を判別した場合にはステップＳ１０に進み、ローカル側のキャッシュメモリ４８－１に対してのみライトデータを転送して書き込み、ステップＳ１１でライト終了応答を返す。例えば図１２に示すように、ホストコンピュータ１０からコントローラ１２－１に対しライト要求が行われたときには、自己のキャッシュ管理テーブルを更新して格納領域を確保し、ライトデータ②を不揮発性キャッシュメモリ４８－１に転送して書き込む。こ

の段階ではディスク装置３０に対する書込みは行わず、その後のライトバック処理で書き込む。

【０１０７】図１０のフローチャートは性能重視モードにおけるライトバック処理を示す。このライトバック処理にあっては、ステップＳ１でアイドル状態を判別すると、ステップＳ２でライトバック条件の成立の有無をチェックし、ライトバックがスケジュールされた条件成立であるとステップＳ５に進み、図１２の③に示すように、ローカル側の不揮発性キャッシュメモリ４８－１からライトバック対象データを読み出してディスク装置３０に書き込み、ステップＳ４でライトバックが済んだキャッシュブロックをキャッシュ管理テーブルから削除するキャッシュ無効化を行う。

５．平均モードのキャッシュ制御

この平均モードで使用する各系統のキャッシュメモリは、共に不揮発性キャッシュメモリとする必要がある。平均モードは信頼性重視モードと性能重視モードの混合型となる。

【０１０８】概略的には、リード要求に対するミスヒット時のステージングはローカル側のキャッシュメモリに対してのみ行い、これは性能重視型と同じである。またライト要求に対するライトデータの書込みはローカル側およびリモート側の両方のキャッシュメモリに対して行い、キャッシュメモリに同一データを格納した二重化を図る。これは性能重視モードと同じである。

【０１０９】このような信頼性重視モードと性能重視モードの混合型である平均モードのキャッシュ制御にあっては、ライト要求が少なくリード要求の多い環境では、キャッシュメモリの容量は性能重視モードと同様に２台のキャッシュメモリを設けたことで２倍となる。またライト要求が多く発行されリード要求の少ない環境では、信頼性重視モードと同様に片側データの送出時でもデータを保障することができる。

【０１１０】図１３のフローチャートは平均モードにおける処理動作を示している。まずステップＳ１でホストコンピュータ１０からのアクセス要求があるとステップＳ２でアクセスモードを判別し、リード要求であればステップＳ３で自己のキャッシュメモリ４８－１、即ちローカル側のキャッシュメモリ４８－１に要求データが存在するか否かのヒット判定を行う。

【０１１１】キャッシュヒットが得られると、ステップＳ４で要求データを読み出してホストコンピュータ１０に転送する。ローカル側のキャッシュメモリ４８－１がミスヒットとなった場合には、ステップＳ５でリモート側のコントローラ１２－２へ共用バス１８を経由してキャッシュ処理を依頼する。このキャッシュ処理の依頼を受けたリモート側のコントローラ１２－２より、リモート側のキャッシュメモリ４８－２でのキャッシュヒットおよび要求データのアドレスを示す応答を受けると、リモート側のキャッシュメモリ４８－２より要求データを読み出して、ステップＳ８でホストへのデータ転送を行う。

【０１１２】一方、ステップＳ６でリモート側よりミスヒットの応答を受けた場合には、ステップＳ７でディスク装置３０からローカル側のキャッシュメモリ４８－１に対するステージングを行った後、ステップＳ８でホストコンピュータ１０へデータを転送する。これら一連のリード処理を終了すると、ステップＳ９でリード終了応答を返す。

【０１１３】一方、ステップＳ２でアクセス要求を受けたローカル側のコントローラ１２－２がライト要求を判別した場合には、ステップＳ１０でリモート側のコントローラ１２－２に共用バスを経由してメッセージを送ってライトデータの転送を通知する。この通知を受けて、リモート側のコントローラ１２－２はキャッシュ管理テーブルを更新してライトデータを書き込むための領域を確保し、ローカル側にレディ応答を返す。

【０１１４】ステップＳ１０でリモート側よりレディ応答があると、ステップＳ１２でローカル側のコントローラ１２－１は自己のキャッシュメモリ４８－１、即ちローカル側のキャッシュメモリ４８－１にライトデータを転送して書き込み、またステップＳ３でリモート側のキャッシュメモリに共用バスを経由してライトデータを転送して書き込ませる。

【０１１５】最終的にローカル側およびリモート側のキャッシュメモリ４８－１，４８－２１に対する書込みが終了すると、ステップＳ１２でライト終了応答を返す。このようなライト処理によりローカル側およびリモート側のキャッシュメモリ４８－１，４８－２に同一データが書き込まれることになる。図１４は平均モードにおけるライトバック処理を示したもので、図６に示した信頼性重視モードのライトバック処理と同じである。

【０１１６】即ち、ライトバック対象となったローカル側のキャッシュメモリ４８－１のダーティデータをディスク装置３０に書き込むライトバック処理を行い、ライトバックが済んでからローカル側のキャッシュメモリ４８－１からライトバック済みデータを削除するキャッシュ無効化を行い、更に共用バス１８でリモート側にメッセージを送って、リモート側のキャッシュメモリ４８－２ライトバック済みのデータを削除するキャッシュ無効化を行わせる。

６．第２発明の全体機能

図１５は図２に示したディスクアレイを対象とした第２発明によるキャッシュ制御の機能を示し、図２のコントローラ１２－１側を代表して示している。

【０１１７】コントローラ１２－１にはキャッシュ制御部４６、不揮発性キャッシュメモリを用いたキャッシュメモリ４８、キャッシュ管理テーブル５０およびディスクアレイ制御部５２が設けられている。キャッシュ制御部４６にはリードキャッシュ制御部４６－１，ライトキ

ャッシュ制御部４６－２およびライトバック制御部４５の各機能が設けられる。

【０１１８】キャッシュ管理テーブル５０はハッシュテーブル６０，ＬＲＵテーブル６２および空スペース管理テーブル６４で構成される。ディスクアレイ制御部５２には、この実施例にあってはＲＡＩＤ０，ＲＡＩＤ１，ＲＡＩＤ３およびＲＡＩＤの形態をもつ動作機能が予め設けられている。ＲＡＩＤの動作形態の指定はホストコンピュータ１０からの指示で論理的に行うことができる。

【０１１９】ディスクアレイ２８は図２に示したハードウェア構成に対応するものであり、縦方向の並びをランクと読んでランク番号Ｒ０～Ｒ３で示し、横方向の並びをポートと読んでポート番号Ｐ０～Ｐ５で示している。図１６は図１５のキャッシュメモリ４８の構造を示す。キャッシュメモリ４８はデータアクセスの最小単位であるキャッシュブロック５４から構成されている。キャッシュブロック５４のサイズは、右側に取り出して示すようにホストコンピュータ１０がアクセスする最小単位であるホストブロック（論理ブロック）の整数倍となる。

【０１２０】図１６にあっては、ホストブロック５４の４倍にキャッシュブロック５４を設定した場合を示している。このキャッシュブロック５４のサイズ設定は通常、システム側で行われ、ユーザは介在しない。図１７はディスク装置３０のデータブロックに対するホストブロックとの関係を示している。ディスクアレイシステムにおけるディスク装置３０のフォーマットは固定長ブロック形式（ＦＢＡ）を採用しており、デコード長が可変のカウントキーデータ形式（ＣＫＤ）はデータのストライピングやパリティの生成が複雑になることから採用しない。

【０１２１】磁気ディスク３０の中に示す両側のインデックスで挟まれた各セクタを構成する固定長データブロック５８は、通常、５１２バイトとなる。ホストブロック５６はディスク装置３０の固定長データブロック５８の整数倍となり、この例ではホストブロック５６はデータブロック５８の４倍であることから２，０４８バイトとなる。

【０１２２】図１８は図１５のキャッシュ管理テーブル５０に設けたハッシュテーブル６０の構造を示す。ハッシュテーブル６０はハッシュエントリ６６とハッシュエントリで指定される登録データ６８で構成される。ハッシュエントリ６６はホストコンピュータ１０からのアクセス要求で得られたデバイス情報をハッシュパラメータとして計算したエントリ値（ハッシュアドレス）で指定され、指定されたハッシュエントリにキャッシュメモリ上に存在するキャッシュブロックＣＢの登録状態が登録データ６８として図示のように登録されている。

【０１２３】このハッシュテーブル６０はホストコンピュータから要求のあったホストブロックがキャッシュメモリ上に存在しているかどうかの判定、即ちヒット／ミスヒット判定に主に使用される。図１９は図１５のキャッシュ管理テーブル５０に設けたＬＲＵテーブル６２を示している。ＬＲＵテーブル６２はキャッシュメモリ４８上に格納されたキャッシュブロックをアクセス順にリンクしている。

【０１２４】例えば新たなキャッシュブロックの書込みが行われると、ＬＲＵテーブル６２の先頭に最新に使用されたキャッシュブロックとしてリンクする。またキャッシュブロック上に存在する任意のキャッシュブロックがリード要求に対し読み出された場合にも、最新に使用されたキャッシュブロックとしてＬＲＵテーブル６２の先頭にリンクされる。

【０１２５】このようなＬＲＵテーブル６２はキャッシュメモリ４８上に存在するデータの追出しと、ディスク装置にキャッシュメモリ上のデータを書き込むライトバック制御に使用される。即ち、ＬＲＵテーブル６２の最後にキャッシュメモリ４８上に存在している時間が最も長い、長時間アクセスされていないキャッシュブロックがリンクされていることから、この最後にリンクされたキャッシュブロックが次に追い出されるキャッシュブロックの候補となる。

【０１２６】また追出し候補となったキャッシュブロックがディスク装置への書込みが済んでいないデータ所謂ダーティデータであった場合には、ディスク装置に書き戻すライトバック処理がスケジュールされ、ライトバック後に追出しを行うことになる。

７．キャッシュ空スペースの確保

図２０はキャッシュメモリ４８におけるキャッシュブロックの使用状態の一例を示している。

【０１２７】キャッシュメモリ４８において、キャッシュブロックＲはライトバックによりディスク装置に同一データが格納されているブロックを示す。キャッシュブロックＷはキャッシュメモリ４８上で書込処理は行われたがライトバックが済んでおらず、ディスク装置に同一データが存在しないダーティデータが存在するブロックを示す。また「空」は未使用のキャッシュブロックである。

【０１２８】ディスクキャッシュ制御にあっては、ホストコンピュータ１０からキャッシュメモリ４８上に存在しないデータに対してのリード要求またはライト要求が指示された場合、未使用の空間から新たなキャッシュブロックを割り当てて、リード要求に伴ってディスク装置３０からステージングされたデータあるいはホストコンピュータから受領したライトデータを書き込む。

【０１２９】図２０のキャッシュメモリ４８における未使用の空キャッシュブロックの空間は、図２１に示す空スペース管理テーブル６４により管理される。空スペース管理テーブル６４は空キャッシュブロックの数をエントリ番号１，２，３，・・・Ｎとし、各エントリ番号ご

とに連続する空ブロックの先頭のブロック情報を登録している。なお、テーブル内容は説明の都合上示したもので、実際には設けられない。

【０１３０】ここで図２０のキャッシュブロック４８におけるＸ方向のアドレスをＸ＝００～０７、Ｙ方向のアドレスをＹ＝００～ｎで表すと、空スペース管理テーブル６４における先頭ブロックのアドレスは「ＣＢＸＹ」で示される。例えば空スペース管理テーブル６４の空ブロック数１個を示すエントリ１には、図２０のキャッシュメモリ４８におけるブロックアドレスＣＢ0103，ＣＢ0004の２つの空ブロックが登録されている。空ブロック２個分を示すエントリ番号２については、２個空ブロックが連続した先頭ブロックのアドレスＣＢ0102が登録されている。

【０１３１】このような空スペース管理テーブル６４を設けることで、新たなキャッシュブロックの割当要求に対してはなるべく物理的に連続したキャッシュブロックを割り当てている。更に空スペース管理テーブル６４で管理されるキャッシュメモリ４８の空スペースは、常に一定量が残るように管理される。

【０１３２】即ち、キャッシュメモリ４８の空スペースの残り容量が規定値以下になるような状況が発生すると、図１９に示したＬＲＵテーブル６２を参照してキャッシュメモリ上の有効データを追い出すためのスケジュールが行われる。この追出し対象となったキャッシュブロックにつき、ディスク装置にも同一データが存在するキャッシュブロックＲの部分については、図１８に示したハッシュテーブル６０からの登録を抹消するキャッシュ無効化を行う。一方、ディスク装置にデータが存在しないダーティデータのキャッシュブロックＷの部分については、ディスク装置へのライトバックがスケジュールされる。

【０１３３】このようなキャッシュメモリ４８の残り容量を規定値以上に保つ処理により、常にキャッシュメモリ上に規定の空き空間が準備されており、従って新たなデータ書込みを伴うリード要求およびライト要求に対し、直ちに処理を開始することができる。

８．ＲＡＩＤ動作形態の制御

図１５に示したディスクアレイ制御部５２はホストコンピュータ１０からの論理デバイスの指定に基づき、ＲＡＩＤ０，１，３または５のいずれかの形態のアクセス動作をディスクアレイ２８に対し実行する。

【０１３４】図２２はランク数４，ポート数６のディスクアレイ２８に対するホストコンピュータが指定する論理デバイス番号と、ＲＡＩＤ動作のためのディスク装置の形態を示している。まずランクＲ０に設けられた６台のディスク装置の内、ポートＰｏ～Ｐ４の５台がホストコンピュータ１０における論理デバイス番号０で指定され、ＲＡＩＤ３の動作形態をとる。この内、ポートＰ０～Ｐ３のディスク装置はデータ用であり、ポートＰ４のディスク装置がパリティ用に固定的に決められる。

【０１３５】尚、ランクＲ０のポートＰ５のディスク装置は予備機として設けられたホットスペアＨＳに割り当てられる。次のランクＲ１のポートＰ０～Ｐ５の５台のディスク装置はホストコンピュータ１０における論理デバイス番号１で指定され、ＲＡＩＤ５の動作形態をとる。ＲＡＩＤ５の動作形態にあっては、ポートＰ０～Ｐ４の中の４台のディスク装置がデータ用、１台がパリティ用となる。パリティ用のディスク装置はセクタ位置が替わるごとに循環する。

【０１３６】ランクＲ２のポートＰ０，Ｐ１およびポートＰ２，Ｐ３に設けられた２台を１組としたディスク装置は、ホストコンピュータ１０における論理デバイス番号２，３で各々指定され、ＲＡＩＤ１の動作形態、即ち２台のディスク装置に同一データを格納したミラーディスクとしての動作形態をもつ。更にランクＲ３のポートＰ０～Ｐ３に設けた４台のディスク装置はホストコンピュータ１０における論理デバイス番号４で指定され、ＲＡＩＤ０の動作形態をとる。即ち、データをストライピングして並列的に書き込む動作は論理デバイス番号０のＲＡＩＤ３の動作形態と同じであるが、冗長性を確保するためのパリティディスクを備えていない。

【０１３７】図２３はＲＡＩＤ３の動作形態におけるキャッシュブロック，ストライピングおよびディスク格納状態を示している。まずキャッシュブロック５４は、この例では４つのホストブロックＬ０～Ｌ３で構成されている。ＲＡＩＤ３の動作モードでディスク装置３０－００～３０－０４に書き込むために、例えばホストブロックＬ０について取り出して示すように、所定数の固定長データブロックのサイズで１６分割したバイトデータＬ０－１～Ｌ０－１６にストライピングする。

【０１３８】このようなストライピングデータについて、データ用のディスク装置３０－００～３０－０３の台数４台に対応したストライピングデータ単位にパリティを計算する。そして４つのストライピングデータに計算したパリティを加えた５つのバイトデータを、５台の磁気ディスク装置３０－００～３０－０４に並列的に書き込む。

【０１３９】図２４はＲＡＩＤ５の動作形態におけるキャッシュブロック，データストライピングおよびディスクアレイ格納状態を示している。この例では説明を簡単にするため、ディスクアレイ２８に並列書込みされるホストブロック数にキャッシュブロックのサイズを一致させている。キャッシュメモリ４８上には並列書込み可能なデバイス、即ちディスク装置３０－００～３０－０３の４台に対応したホストブロック数４のサイズをもつキャッシュブロック５４－１～５４－４が設けられ、図示のように連続するキャッシュブロック５４－１～５４－４に、同じく連続するホストブロックＬ０～Ｌ１５が格納されていたとする。

【０１４０】このような場合には、例えばキャッシュブロック５４－１を取り出し、キャッシュブロック５４－１を構成する４つのホストブロックＬ０～Ｌ３からパリティを計算し、ホストブロックＬ０～Ｌ３にパリティＰ１を加えた５つのデータを並列的にディスク装置３０－００～３０－０４に書き込む。通常、ＲＡＩＤ５の動作形態にあっては、データ長の短いホストブロック１個単位のアクセスが行われる。この通常のＲＡＩＤ５の動作形態における書込みは、例えば図示のようにディスクアレイ２８側にホストブロックＬ０～Ｌ３およびパリティＰ１が格納されている状態で、新たに論理ブロックＬ０のみを書き込む場合には、次のようなリードモディファイライトを実行する。

【０１４１】まずディスク装置３０－００および３０－０４により旧ブロックＬ０と旧パリティＰ１を読み出す。これを（Ｌ０）old，（Ｐ１）old とする。続いて中間パリティ（Ｐ１）int を

（Ｐ１）int ＝（Ｐ１）old ＋（Ｌ０）old

として、排他論理和を求める。続いて中間パリティ（Ｐ１）int と新ブロック（Ｌ０）new から新パリティ（Ｐ１）new を

（Ｐ１）new ＝（Ｐ１）int ＋（Ｌ１）new

として求める。新ブロック（Ｌ０）new と計算で求めた新パリティ（Ｐ１）newをディスク装置３０－００，３０－０４に対し並列的に書き込む。

【０１４２】このようなＲＡＩＤ５の動作形態における並列書込み可能なディスク台数以下のホストブロックの書込みについては、旧パリティと旧データの読出しを伴うことからディスク２．５回転分の処理時間がかかり、これがライトペナルティとなる。第２発明にあっては、ＲＡＩＤ５の動作形態にあっても可能な限りＲＡＩＤ３の動作形態によるディスクアレイに対するライトアクセスを行って、ライトペナルティの低減を図る。

【０１４３】図２５はＲＡＩＤ１の動作形態におけるキャッシュブロックとディスク格納状態を示している。ＲＡＩＤ１は２台のディスク装置３０－２０，３０－２１に同一データを書き込むミラーディスクとしての使い方であり、キャッシュブロック５４からホストブロックＬ０～Ｌ３ごとに順次取り出してディスク装置３０－２０，３０－２１のそれぞれに並列的に書き込む。

【０１４４】図２６はパリティディスクをもたないＲＡＩＤ０の動作形態のキャッシュブロック，ストライピングおよびディスク格納状態を示す。尚、図２４のＲＡＩＤ５の場合と同様、説明を簡単にするため、４台の磁気ディスク装置３０－３０～３０－３４に並列書込みする論理ブロック数４に一致するキャッシュブロックのサイズとした場合を例にとっている。

【０１４５】このＲＡＩＤ０の動作形態では、キャッシュブロック５４を構成する４つのホストブロックＬ０～Ｌ３を、４台のディスク装置３０－３０～３０－３４のそれぞれに割り当てて並列的に書き込み、この場合にパリティディスクはもたないことから、パリティの計算は行わない。

９．ステージング処理の高速化

ディスクキャッシュ制御にあっては、ホストコンピュータ１０からのリード要求に対しキャッシュメモリ４８上に要求データが存在せずにミスヒットとなった場合には、ディスク装置から要求データを読み出してキャッシュメモリ上に書き込むステージングを行う。

【０１４６】本発明にあっては、ＲＡＩＤ０，１，３および５のいずれの動作形態にあっても、ディスクアレイ２８の同時アクセス可能な全ディスク装置からの並列リードによるステージングで要求データのキャッシュブロックに対するステージングを高速で行う。図２８はＲＡＩＤ５の動作形態を例にとった本発明のステージングを示している。いま１６ホストブロックのサイズをもつキャッシュブロック５４－１がキャッシュメモリ４８上に展開されており、このキャッシュブロック５４－１のうちの最後の２つのホストブロックＬ１４，Ｌ１５のみが有効データであり、またホストブロックＬ１４，Ｌ１５はディスク装置への書込みが済んでいないダーティデータであったとする。

【０１４７】ここで、ホストコンピュータ１０より論理ブロックＬ４～Ｌ７を指定したリード要求①がホストコンピュータから行われたとする。このリード要求に対し、キャッシュメモリ上に要求データが存在しないことからミスヒットとなり、ステージングを行う。ステージングはキャッシュブロック５４－２に示すように、ディスクアレイ２８のディスク装置３０－１０～３０－１２，３０－１４の４台をアクセスしてホストブロックＬ４～Ｌ７をリードしてキャッシュブロック５４－２上に書き込み、ステージングが完了するとヒット状態となることから、要求データをホストコンピュータ１０に転送する。

【０１４８】このようなステージングが終了した後に本発明にあっては、それ以降のアクセス要求に対するキャッシュメモリのヒット効率を向上してキャッシュ効果を最大限に発揮するため、プリフェッチ動作を行う。即ち、キャッシュブロック５４－３に示すように、ステージングが済んだホストブロックＬ４～Ｌ７に後続する並列アクセス可能なディスク数４に一致する４つのホストブロックＬ８～Ｌ１１をディスク装置３０－１０，３０－１１，３０－１３，３０－１４から読み出してプリフェッチする。

【０１４９】ＲＡＩＤ５にあっては、基本的にはリード転送速度はデバイス速度以上は出ることがなく、またステージング速度がＲＡＩＤ３より遅く、少量データのリード処理が中心になることから、プリフェッチ量は次のパリティグループまでとする。ここで、パリティグループとはパリティディスクを同一とする各ディスク装置の

データブロックを意味する。例えば、図２７（Ｅ）におけるデータブロックＬ０～Ｌ３，Ｌ４～Ｌ７などのことを意味する。

【０１５０】このようにＲＡＩＤ５にあっては、プリフェッチ量は次のパリティグループまでとなるが、キャッシュブロックサイズをパリティグループのデータブロックサイズと同一になるようにストライピングを行えば、即ち図２４に示したストライピングを行えばキャッシュブロックの中にステージングしたデータブロックに続くキャッシュブロックの最後までのデータブロックがプリフェッチ量となり、このプリフェッチをステージングに含めてＲＡＩＤ３的なステージングが可能となる。

【０１５１】一方、ＲＡＩＤでは図２３に示したように、１つのホストブロックが各ディスク装置に分散して格納されており、ステージングが高速にできること、および大容量のデータが要求される場合が多いことを考慮し、アクセス要求が行われたステージングデータを含むキャッシュブロックの最後までのデータブロックをプリフェッチする。

【０１５２】再び図２７を参照するに、キャッシュブロック５４－３に召すステージングしたホストブロックＬ４～Ｌ７に続くブロックＬ８～Ｌ１１をプリフェッチした状態にあっても、キャッシュブロック上の有効データは２箇所に分散している。このようなステージングに伴うプリフェッチ後の状態でキャッシュブロック上の有効データが分散している場合には、キャッシュブロック５４－４に示すようにホストブロックＬ１２，Ｌ１３に対応するデータブロックをディスクアレイ２８のディスク装置３０－１２，３０－１３から読み出して内部的にステージングし、キャッシュブロック上で有効データを連続データとする。

【０１５３】このようにキャッシュブロック上の有効データを連続データとすることで、キャッシュブロック上での以後のデータ管理を容易にする。例えば、キャッシュブロック上で有効データが複数箇所に分散して存在していたとすると、最終的には分散した各ホストブロックに対応したマップ管理が要求される。これに対し、キャッシュブロック上に連続した領域を有効データとすることで、有効データの一部が書き替えられた際のライトバック処理において、ＲＡＩＤ３的な並列書込みによる高速化を図ることができる。

【０１５４】１０．大量シーケンシャルデータの追出し促進

一般にキャッシュ制御にあっては、キャッシュメモリ４８上の同一データを頻繁にリードするとキャッシュ効果を発揮することができる。これに対し、ホストコンピュータ１０がキャッシュメモリ４８上に存在しない大量データのリード要求を行ったとき、ミスヒットに伴ってディスクアレイ２８から要求された大量データをキャッシュメモリ４８上にステージングしてホストコンピュータ１０に転送するようになる。

【０１５５】このようにステージングされた大量のデータがその後、参照されないにも関わらず、そのままキャッシュメモリ４８上に残されていると、キャッシュスペースが有効に利用されず、ヒット効率が低下することになる。しかしながらＬＲＵテーブル６２を用いたキャッシュメモリ４８の管理では、一度しか参照されなかった大量データであっても直ちに無効化することはできない。

【０１５６】そこで本発明のディスクキャッシュ装置にあっては、ホストコンピュータ１０からの一定量を越えるデータのリード要求に伴って大量データのステージングが行われた場合には、本来、ＬＲＵテーブル６２の先頭にステージングされたキャッシュブロックを登録すべきところ、ＬＲＵテーブル６２の終端側に近い位置にステージングされた大量データのキャッシュブロックを登録する。

【０１５７】図２９のフローチャートは大量データのステージング処理を示す。まずステップＳ１でホストコンピュータ１０からのリード要求に伴うミスヒットでディスクアレイ２８側からのステージングが行われたならば、ステップＳ２で、ステージングされたデータ量が所定値例えば１００キャッシュブロック以上か否かチェックする。

【０１５８】１００キャッシュブロック未満であれば通常通り、ステップＳ３でＬＲＵテーブル６２の先頭に登録する。一方、１００キャッシュブロック以上であった場合には、ステップＳ４でＬＲＵテーブル６２の終端側の近くに登録する。このようなステージングされた大量データについて、ＬＲＵテーブル６２の終端側に強制的に登録することで、一度しか参照されないステージングされた大量データは比較的短い時間の経過後にハッシュテーブル６０から削除されて追い出されることになり、キャッシュスペースが不要に占有されてしまうことを防止できる。

１１．パリティデータのステージング

本発明のディスクキャッシュ装置におけるＲＡＩＤ５の動作状態において、原則的にはキャッシュメモリ４８上にパリティデータのステージングは行っていない。しかしながらＲＡＩＤ５の動作状態にあっては、ディスクアレイに対する書込時に新パリティを計算するためにディスクアレイ２８から旧データおよび旧パリティを読み出す処理を必要とするため、処理時間が長くなるライトペナルティをもつ。

【０１５９】そこで本発明のステージング処理の他の実施例にあっては、オプション指定によりパリティデータのステージングを可能とする。図３０のフローチャートはパリティデータのステージングを可能とする本発明のステージング処理の他の実施例を示す。このステージング処理にあっては、ホストコンピュータ１０からのリー

ド要求に対するキャッシュメモリ４８のミスヒットでステージング処理が起動されると、まずステップＳ１で動作形態がＲＡＩＤであり且つパリティデータのステージングがオプション指定されているか否かチェックする。

【０１６０】パリティデータのステージングを行わせるオプション指定は、ホストコンピュータ１０からのコマンドあるいはメンテナンス端末からの指示によりセットされる。ステップＳ１でパリティデータのステージングを示すオプション指定が有効であるとステップＳ２に進み、パリティデータを含むデータをディスクアレイ２８からキャッシュメモリ４８上にステージングする。ステップＳ１でオプション指定がなければステップＳ３に進み、通常のパリティを含まないデータのディスクアレイ２８からのステージングを行う。

【０１６１】図３１はパリティデータをステージングしない場合とステージングした場合のライトバック処理を対比して示す。図３１（Ａ）はパリティデータをステージングしていない場合であり、キャッシュブロック５４－１０のホストブロックＬ１のみが有効データであったとする。

【０１６２】このキャッシュブロック５４－１０の有効データであるホストブロックＬ１をディスクアレイ２８にライトバックする場合には、図２４のＲＡＩＤ５におけるディスク書込みに示したようにディスク装置３０－１１から対応する旧データを読み出し、またディスク装置３０－１４から旧パリティを読み出し、中間パリティを生成した後に新データから新パリティを作成し、最終的にディスク装置３０－１１，３０－１４に新データと新パリティを書き込む処理を必要とする。

【０１６３】このため、キャッシュブロック５４－１０上の有効ブロックＬ１をディスクアレイ２８に書き込むためにはディスク装置の２．５回転分の時間が必要となる。これに対し図３１（Ｂ）に示すキャッシュブロック５４－１０上にパリティデータＰ１をステージングしていた場合には、有効ブロックＬ１の更新に際しキャッシュブロック５４－２０上で新たなパリティデータＰ１を計算でき、ディスクアレイ２８側をリードアクセスすることなく直接、更新ブロックＬ１と新パリティデータＰ１をディスク装置３０－１１，３０－１４に書き込むことができる。

【０１６４】このため、ライトバック処理はディスク装置１．５回転分の時間で済み、通常のＲＡＩＤ５の動作状態のディスク書込み処理におけるライトペナルティは発生しない。勿論、パリティデータをキャッシュメモリ４８にステージングした場合には、その分だけキャッシュメモリ４８の有効利用領域が低減することになる。

１２．ライトバックデータの並び替えによる最適化

図２２の空スペース管理テーブル６４で説明したように、キャッシュメモリ４８上の空スペースが一定値以下になった場合は、図１９に示すＬＲＵテーブル６２を参照し、使用されずに最も長い期間格納されているキャッシュブロックを追い出すためのライトバック処理がスケジュールされる。

【０１６５】図３２はライトバック処理がスケジュールされる際のキャッシュブロック６２の一例を示している。通常のキャッシュ制御にあっては、ＬＲＵテーブル６２の最後にリンクされているキャッシュブロックＣＢ－ｆをライトバック対象とするスケジュールを行う。これに対し本発明のキャッシュ制御にあっては、ライトバック条件が成立したとき、即ちキャッシュメモリ４８の空ブロック数が規定の閾値以下となったときには、ＬＲＵテーブル６２の最後のキャッシュブロックから所定ブロック数αだけ遡った範囲をライトバック対象範囲６８として指定するスケジュールを行う。

【０１６６】この場合にはキャッシュブロックＣＢ－ａ～ＣＢ－ｆの６つがライトバック対象範囲６８に含まれるように指定される。ここでライトバック対象範囲６８に含まれる斜線で示すキャッシュブロックＣＢ－ｂ，ＣＢ－ｅ，ＣＢ－ｆの３つが、図３３に示すホストブロックを含んでいたとする。即ち、キャッシュブロックＣＢ－ｂはホストブロックＬ４～Ｌ７の対応データを含み、キャッシュブロックＣＢ－ｅはホストブロックＬ８～Ｌ１１の対応データを含み、キャッシュブロックＣＢ－ｆはホストブロックＬ０～Ｌ３の対応データを含んでいたとする。

【０１６７】このようなホストブロック対応データを見ると、キャッシュブロックＣＢ－ｆ，ＣＢ－ｂ，ＣＢ－ｅの順番に並び替えることで、ホストブロックＬ０～Ｌ１１の対応データを連続データとすることができる。ホストブロック対応データを連続データとするライトバック対象範囲６８に含まれるキャッシュブロックの並び替えを行った場合には、ディスクアレイ２８に各キャッシュブロックをライトバックする際にＲＡＩＤ３的な並列書込みを高速で実行できる。

【０１６８】即ち、キャッシュブロックＣＢ－ｆ，ＣＢ－ｂ，ＣＢ－ｅに含まれる４つの論理ブロック対応データのそれぞれからパリティＰ１，Ｐ２，Ｐ３を計算し、順番に４つのデータとパリティを並列的にディスクアレイ２８のディスク装置３０－１１～３０－１４に連続的に書き込む高速ライトバック処理が達成できる。このようなライトバック対象範囲６８に含まれるキャッシュブロックの論理ブロック対応データを連続するような並び替えによるライトバックが完了すると、図３２に示したＬＲＵテーブル６２は図３４の状態となる。

【０１６９】図３５のフローチャートはキャッシュブロックの並び替えを伴うライトバックのスケジュールが行われるライト処理を示し、図２８のステップＳ１３に示したライト処理の詳細を示すことになる。まずステップＳ１でホストコンピュータ１０からのライト要求に伴って、ライトキャッシュ制御部４６－２でキャッシュメモ

リ４８上に確保された８ブロックに対しライトデータを書き込む。

【０１７０】ライトデータの書込みが済むと、ステップＳ２で図２２に示した空スペース管理テーブル６４を参照してキャッシュメモリ４８の空ブロック数を求め、空ブロック数が予め定めた閾値以下か否かチェックする。閾値以下であればステップＳ３に進み、図３２に示したようにＬＲＵテーブル６２の末尾から遡った所定数αのブロックを含む範囲をライトバック対象範囲６８として指定する。

【０１７１】ライトバック対象範囲６８に含まれるキャッシュブロックがディスクアレイ２８への書込みが済んでいないライトバック対象となるダーティデータか、既にディスクアレイ２８への書込みが済んでいるライトバック対象から除外される非ダーティデータか否かチェックする。もしライトバック対象範囲６８に非ダーティデータのキャッシュブロックが含まれていれば、ステップＳ５で非ダーティデータのキャッシュブロックをハッシュテーブル６０から削除してキャッシュ無効化を行う。

【０１７２】続いてステップＳ６でライトバック対象となるダーティデータのキャッシュブロックの有無をチェックし、ダーティデータのキャッシュブロックが存在すると、ステップＳ７でライトバックをスケジュールする。このライトバックのスケジュールにあっては、図３２，図３３に示したようにライトバック対象範囲６８に含まれるキャッシュブロックについて、もし可能であればホストブロックに対応するデータが連続データとなるようにキャッシュブロックの並び替えを行った後にライトバック対象としてスケジュールする。

【０１７３】尚、ライトバック処理のスケジュールは図３５に示すライト処理の完了時のみならず、図２８に示したリード要求でキャッシュメモリがミスヒットとなったときのディスクアレイからのステージング終了時についても同様にスケジュールされることになる。

１３．ライトバック処理の最適化

図３６はＲＡＩＤ５の動作形態におけるライトバック処理の一例を示している。ここでキャッシュブロック５４のサイズは１６ホストブロックであり、ホストブロックＬ０～Ｌ１５に対応するデータを格納することができ、この内、ホストブロックＬ２～Ｌ９に対応するデータが有効データとして書き込まれ、ライトバック対象データ７５となっている。

【０１７４】一方、ディスクアレイ２８側はディスク装置３０－１０～３０－１４のそれぞれは１ホストブロックのサイズに相当するセクタ単位にデータを読み書きしており、同時並列的には４つのホストブロックをアクセスすることができる。キャッシュブロック５４のライトバック対象データ７５に含まれるホストブロックＬ２～Ｌ９対応データについて、ディスクアレイ２８のストライピング状態に対応し、先頭のホストブロックＬ２，Ｌ３と末尾のホストブロックＬ８，Ｌ９の各対応データについてＲＡＩＤ５モードによる書込みを行う。

【０１７５】即ち、対応する旧データおよびパリティデータをディスクアレイ２８から読み出して、書込みを行う新データを用いて新パリティを計算した後に並列的に書き込むリードモディファイライトを行う。これに対し、ディスクアレイ２８の並列書込み可能な４台のディスク装置に対応する連続する４つのホストブロックＬ４～Ｌ７の対応データについては、ＲＡＩＤ３モードによる書込みを行う。即ち、ホストブロックＬ４～Ｌ７の対応データからパリティＰ２を計算し、直接ディスクアレイ２８に書き込む。

【０１７６】このようなＲＡＩＤ５モードにおけるライトバック処理について、通常のＲＡＩＤ５モードに書込みに加えてＲＡＩＤ３モードによる高速書込みを組み合わせることで、最適化されたライトバック処理を行うことができる。図３７はＲＡＩＤ５の動作状態におけるライトバック処理において、ライトバック対象データ７５の中の連続するホストブロック対応データが、ディスクアレイ２８側の並列書込みディスク台数より僅かに少なかった場合の最適化処理を示している。

【０１７７】いまキャッシュブロック５４の中にライトバック対象データ７５としてホストブロックＬ１～Ｌ１０の対応データが格納されていたとする。中央のホストブロックＬ４～Ｌ７の４つのデータについてはＲＡＩＤ３的な並列アクセスができる。これに対し、左側のホストブロックＬ１～Ｌ３対応データおよび右側のホストブロックＬ８～Ｌ１１対応データは並列アクセスディスク台数４台に対し１つホストブロック対応データが少ない。

【０１７８】このような場合のライトバックにあっては、キャッシュブロック５４－２に示すようにホストブロックＬ０対応データをディスクアレイ２８のディスク装置３０－１０から①に示すようにリードしてステージングし、またホストブロックＬ１１の対応データについてもディスクアレイ２８のディスク装置３０－１４により②に示すようにリードしてステージングし、並列アクセス可能なディスク台数の整数倍となるホストブロック数の連続データとする。

【０１７９】このようなホストブロックＬ０～Ｌ１１の連続対応データがキャッシュブロック５４－２に上に得られれば、ホストブロックＬ０～Ｌ３，ホストブロックＬ４～Ｌ７，ホストブロックＬ８～Ｌ１１の４つずつにストライピングして、それぞれパリティＰ１，Ｐ２，Ｐ３を求めた後、①④⑤に示すように並列的にディスクアレイ２８に書き込むことで、ＲＡＩＤ３的な高速ライトバック処理を行うことができる。

【０１８０】図３８のフローチャートは図３６，図３７に示したＲＡＩＤ５のライトバック処理を含む本発明のキャッシュ制御における全体的なライトバック処理を示

す。まずステップＳ１でコントローラ１２がアイドル状態か否かチェックし、アイドル状態にないときにはステップＳ１４で要求事項の処理、即ち通常の処理を行っている。

【０１８１】アイドル状態が得られるとステップＳ２に進み、ライトバック処理がスケジュール済み、即ちライトバック条件が成立していればステップＳ３に進み、ＲＡＩＤモードをチェックする。ＲＡＩＤ５モードであればステップＳ４に進み、図３２，図３３に示したようなライトバック対象範囲におけるキャッシュブロックの連続ブロックへの並び替えを行う。なお、この連続ブロックへの並び替えは図３５に示したライト処理の段階におけるスケジューリングとして行っておいてもよい。

【０１８２】続いてライトバック対象となった最初のキャッシュブロックを参照し、ディスクアレイ２８側の並列書込み可能なディスク台数に対応した全台数分のホストブロック対応データが有効か否かチェックする。もし全デバイス分のホストブロック対応データが有効であればステップＳ６に進み、ＲＡＩＤ３モードによる書込処理を実行する。即ち、全デバイス分の対応データからパリティを計算した後、データおよびパリティを並列的に同時書込みする。

【０１８３】一方、全デバイス分のキャッシュブロック対応データが有効でなかった場合には、ステップＳ７で不足データは1ディスク分か否かチェックする。不足データが1ディスク分であった場合には、ステップＳ８で、図３７に示したように不足ブロックに対応するデータをディスクアレイからステージングし、ステージング後にステップＳ６に進んでＲＡＩＤモードでの書込処理を行う。

【０１８４】ステップＳ７で不足分が２ディスク分以上であった場合にはステップＳ９に進み、ＲＡＩＤ５モードでの書込処理を行う。即ち、旧データおよび旧パリティをリードして中間パリティを計算し、更に新データから新パリティを計算し、新データおよび新パリティを並列的にディスクアレイに対し同時書込みする。ステップＳ６またはステップＳ９の書込処理が済むと、ステップＳ１０でライトバック対象となっているキャッシュブロックの全ブロックの処理が終了したか否かチェックし、終了していなければステップＳ４に戻り、同様な処理を繰り返し、終了していればステップＳ１で再び待機状態となる。

【０１８５】一方、ステップＳ３でＲＡＩＤ３の動作モードが判別された場合にはステップＳ１１に進み、図２３に示したストライピングおよび書込みとなるＲＡＩＤ３モードの書込処理を行う。またステップＳ３でＲＡＩＤ０モードが判別された場合には、ステップＳ１３で、図２５に示すＲＡＩＤ０モードの書込みを行う。更にステップＳ３でＲＡＩＤ１モードが判別された場合にはステップＳ１３に進み、図２６に示したパリティディスクをもたないＲＡＩＤ１モードの書込みを行う。

１４．統計的処理によるキャッシュブロックサイズの自動設定

本発明を含めたディスクキャッシュ装置を備えた計算機システムにおいて、ユーザが必ず設定を必要とするパラメータはホストブロックサイズ即ち論理ブロックサイズである。

【０１８６】これに対し、キャッシュブロックサイズの設定は通常、システム側で行っている。図１６に示したように、キャッシュブロック５４のサイズはホストブロック５６が複数個集まったブロックとして構成される。このような従来の固定的なキャッシュブロックサイズのシステム側での設定に対し、本発明のディスクキャッシュ装置にあっては、最初のシステム立上げ時はシステム側が固定的にキャッシュブロックサイズを設定するが、次回以降の立上げ時にあってはユーザの使用環境に応じて自動的に最適なキャッシュブロックサイズに変更される。

【０１８７】図３９のフローチャートは最適化されたキャッシュブロックサイズを自動設定するための統計処理を示している。最初のシステム立上げ時にあっては、システム側に固定的に設定したキャッシュブロックサイズによるキャッシュ制御が行われており、この状態でホストアクセスをステップＳ１で判別すると、ステップＳ２でリード要求またはライト要求が行われたホストブロックサイズを記録する。

【０１８８】ホストブロックサイズの記録処理はステップＳ５で規定周期に達するまで繰り返し行われている。ステップＳ３で規定周期例えば1カ月程度の時間に達するとステップＳ４に進み、ステップＳ２で記録されたホストブロックサイズから平均ブロックサイズを計算する。続いてステップＳ５で最適ブロックサイズの設定値を、ステップＳ４で求めた平均ブロックサイズに更新する。この最適キャッシュブロックサイズの更新が済むと、ステップＳ２で記録されたホストブロックサイズはメモリからクリアされ、次の周期の記録に入る。

【０１８９】この結果、システムの運用中に統計的処理で得られたキャッシュブロックサイズが最適ブロックサイズとして保持されることになる。そして一旦、システムの運用を止め、次に電源投入によるシステム立上げを行うと、図３９の統計処理で求められている最適キャッシュブロックサイズに自動的にキャッシュブロックサイズを変更してキャッシュ制御を動作する。

【０１９０】図４０はホストブロックサイズに対し最適なキャッシュブロックサイズの一例を示している。いま並列アクセスされるディスクアレイ２８はディスク装置３０－１０～３０－１３の４台であり、それぞれのレコードにおける1セクタを構成するデータブロック５８は５１２バイトであったとする。

【０１９１】この５１２バイトのデータブロック５８に

対し、ホストブロックサイズが２倍の１，０２４バイトであったとすると、例えばホストブロックＬ０は対応するディスク装置３０－１０の２セクタ分に跨がる。キャッシュブロック５４は４台の並列アクセス可能なディスク装置３０－１０～３０－１３のブロックデータサイズに一致するサイズをもつことがＲＡＩＤ３的なアクセスを可能とするために最適である。

【０１９２】従って、キャッシュブロック５４は４つのホストブロックＬ０～Ｌ３に対応し、１つのホストブロックは１，０２４バイトとなることから、キャッシュブロックサイズは４，０９６バイトとすればよい。尚、上記の実施例は６ポート，４ランクのディスクアレイを例にとるものであったが、ディスクアレイのポート数およびランク数は必要に応じて適宜に定めることができる。

【０１９３】また第１発明の実施例にあっては、入出力デバイスとしてディスクアレイを例にとっているが、アレイ構成をもたない通常のディスク装置についてもそのまま適用することができる。更に第２発明にあっては、ホストコンピュータに対し２系統のキャッシュメモリを有するコントローラを設けた場合を例にとっているが、１系統のコントローラにキャッシュメモリを設けたものであったもよいことは勿論である。

【０１９４】更に本発明は実施例に示された数値による限定は受けない。

【０１９５】

【発明の効果】以上説明してきたように第１発明によれば、必要に応じてキャッシュメモリの信頼性を維持しながらの性能の向上、性能を維持しながらの信頼性の向上、および両者の中間的性能という柔軟な運用を選択的にとることができ、ディスクキャッシュを備えたシステム全体の運用効率を向上させることができる。

【０１９６】また第２発明にあっては、ディスクアレイを対象とした所謂ＲＡＩＤ型の動作特性を活かしながらキャッシュ性能を最大限に発揮することができる。

【図面の簡単な説明】

【図１】本発明の原理説明図

【図２】本発明のハードウェア構成を示した実施例構成図

【図３】図２のコントローラのハードウェア構成を示した実施例構成図

【図４】本発明の第１発明の処理機能を示した説明図

【図５】図４の信頼性重視モードの処理動作を示したフローチャート

【図６】図４の信頼性重視モードにおけるライトバック処理を示したフローチャート

【図７】信頼性重視モードでのリード要求に対しローカル側でミスヒットとなった場合の動作を示した説明図

【図８】信頼性重視モードでのライト要求に対するキャッシュ書込動作の説明図

【図９】図４の性能重視モードの処理動作を示したフローチャート

【図１０】図４の性能重視モードにおけるライトバック処理を示したフローチャート

【図１１】性能重視モードでのリード要求に対しローカル側でミスヒット、リモート側でヒットとなった場合の動作を示した説明図

【図１２】性能重視モードでのライト要求に対するキャッシュ書込動作の説明図

【図１３】図４の平均モードの処理動作を示したフローチャート

【図１４】図４の平均モードにおけるライトバック処理を示したフローチャート

【図１５】ディスクアレイを対象とした第２発明の処理機能を示した説明図

【図１６】キャッシュメモリのブロック分けとホストブロックの関係を示した説明図

【図１７】ホストブロックとデータブロックとの関係を示した説明図

【図１８】図１５のハッシュテーブルの説明図

【図１９】図１５のＬＲＵテーブルの説明図

【図２０】図１５のキャッシュメモリの使用状態の一例を示した説明図

【図２１】図１５の空スペース管理テーブルの説明図

【図２２】図１５のディスクアレイに対するＲＡＩＤ０，１，３，５のディスク割当てを示す論理デバイスの説明図

【図２３】ＲＡＩＤ３におけるキャッシュブロック、ストライピング及びディスク格納状態を示した説明図

【図２４】ＲＡＩＤ５におけるキャッシュブロック、ストライピング、ディスク格納状態を示した説明図

【図２５】ＲＡＩＤ１におけるキャッシュブロックとディスク格納状態を示した説明図

【図２６】ＲＡＩＤ０におけるキャッシュブロックとディスク格納状態を示した説明図

【図２７】ステージング処理に伴なうプリフェッチ動作を示した説明図

【図２８】プリフェッチのステージングを含む図１５のリード処理を示したフローチャート

【図２９】大量シーケンシャルデータのステージング処理を示したフローチャート

【図３０】パリティを含むステージング処理を示したフローチャート

【図３１】パリティを含まない場合と含む場合のステージングに対応するライトバック処理の相違を示した説明図

【図３２】ＬＲＵテーブルにおけるライトバック対象範囲の説明図

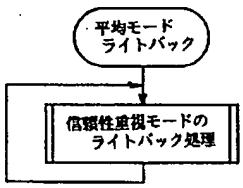
【図３３】ライトバック対象範囲におけるキャッシュブロック並び替えとライトバック動作を示した説明図

【図３４】ライトバック完了後のＬＲＵテーブルの説明

図
【図３５】図２７のライト処理の詳細を示したフローチャート
【図３６】ＲＡＩＤ５におけるライトバック動作の説明図
【図３７】ＲＡＩＤ５において不足ブロックをキャッシュブロックにステージングするライトバック動作の説明図
【図３８】ライト完了後に行うライトバック処理を示したフローチャート
【図３９】キャッシュブロックサイズの求める統計処理のフローチャート
【図４０】最適なキャッシュブロックサイズの決め方の一例を示した説明図
【図４１】揮発性メモリを用いた従来のディスクキャッシュ装置の説明図
【図４２】図４１の従来装置のキャッシュヒット時のリード処理を示した説明図
【図４３】図４１の従来装置のキャッシュミスヒット時のリード処理を示した説明図
【図４４】図４１の従来装置のライト処理を示した説明図
【図４５】揮発性メモリと不揮発性メモリを混在させた従来装置の説明図
【図４６】図４５の従来装置のキャッシュミスヒット時のリード処理を示した説明図
【図４７】図４５の従来装置のライト処理を示した説明図
【符号の説明】
１０：ホストコンピュータ（上位装置）
１２，１２－１，１２－２：コントローラ（ディスク制御手段）
１４－１，１４－２：チャネル装置
１６：チャネルインタフェース（ＳＣＳＩ）
１８，１８－１，１８－２：共用バス（共用バス手段）
２０：ブリッジ回路
２２－１，２２－２：サブコントローラ
２４－１～２４－６，２６－１～２６－６：アダプタ
２８：ディスクアレイ
３０，３０－００～３０－３５：ディスク装置
３２：ＣＰＵ
３４：ＲＯＭ
３６：ＤＲＡＭ
３８：上位インタフェース部
４０：ＳＣＳＩ回路部
４２：バスインタフェース部
４４：内部バス
４５：ライトバック処理部
４６：キャッシュ制御部
４６－１：リードキャッシュ制御部
４６－２：ライトキャッシュ制御部
４８：キャッシュメモリ
４８－１，４８－２：不揮発性キャッシュメモリ
５０：キャッシュ管理テーブル
５２：ディスクアレイ制御部
５４，５４－１～５４－４：キャッシュブロック
５６：ホストブロック
５８：デバイスブロック（１セクタ）
６０：ハッシュテーブル
６２：ＬＲＵテーブル
６４：空スペース管理テーブル
６６：ハッシュエントリ
６８：登録データ
７０，７０－１，７０－２：動作モード設定部
７２，７２－１，７２－２：信頼性重視モードキャッシュ制御部
７４，７４－１，７４－２：性能重視モードキャッシュ制御部
７５：ライトバック対象範囲
７６，７６－１，７６－２：平均モードキャッシュ制御部

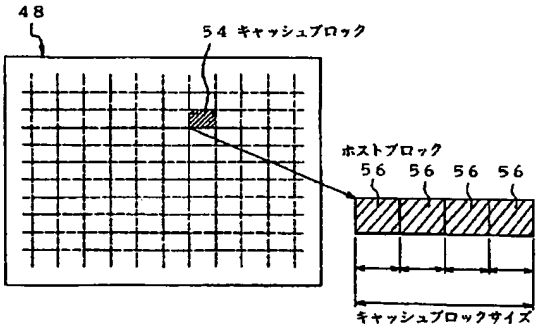
【図１４】

図4の平均モードにおけるライトバック処理を示したフローチャート



【図１６】

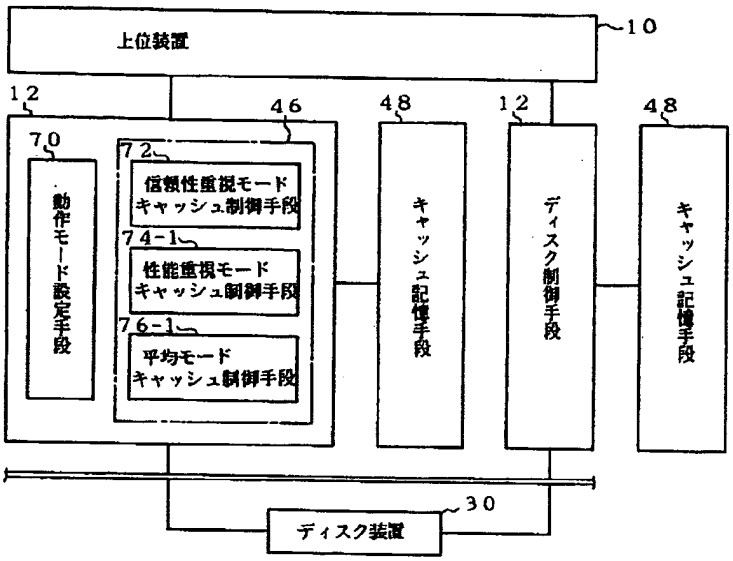
キャッシュメモリのブロック分けとホストブロックの関係を示した説明図



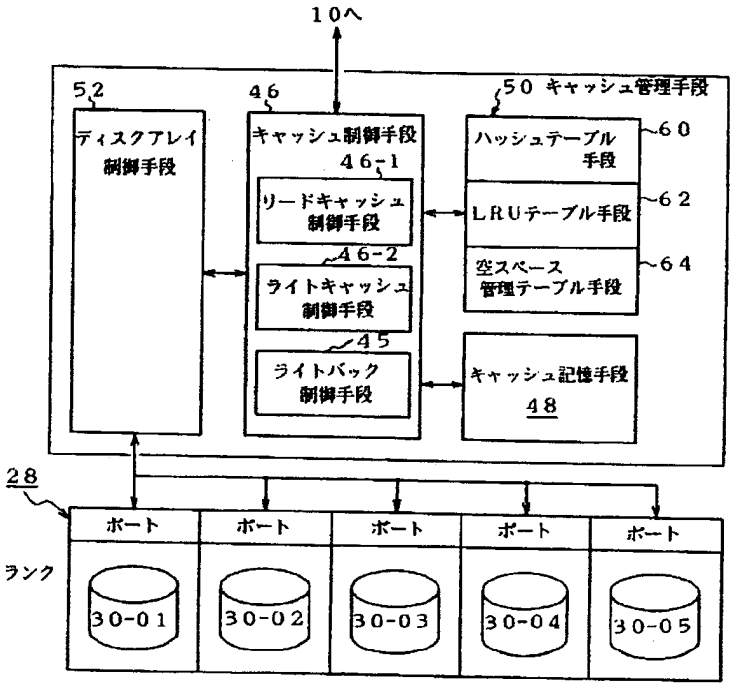
【図１】

本発明の原理説明図

（Ａ）第１発明

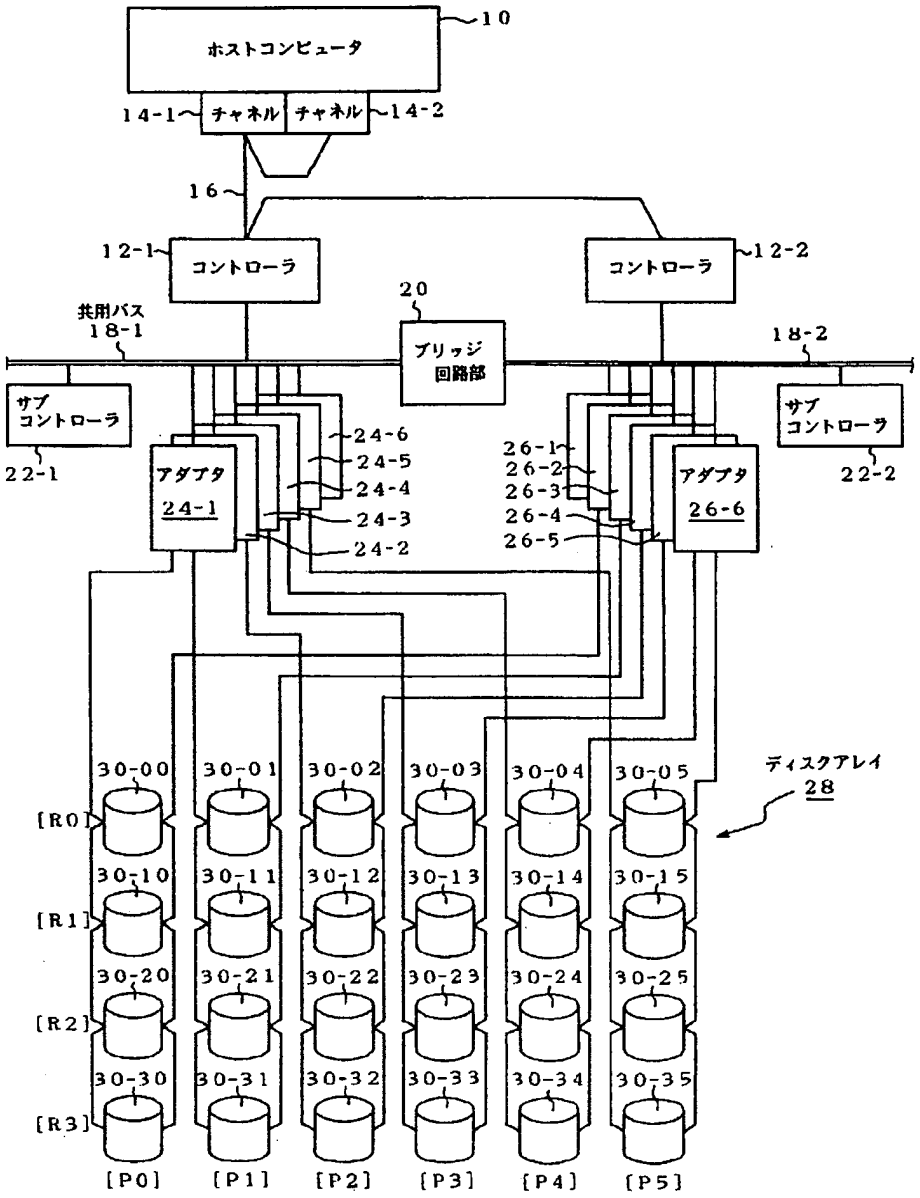


（Ｂ）第２発明



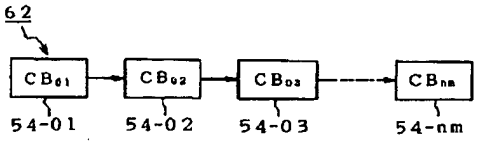
【図２】

本発明のハードウェア構成を示した実施例構成図



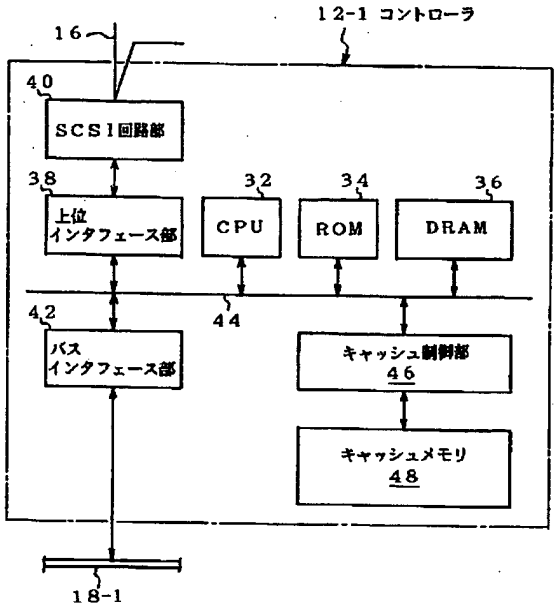
【図１９】

図１５のＬＲＵテーブルの説明図



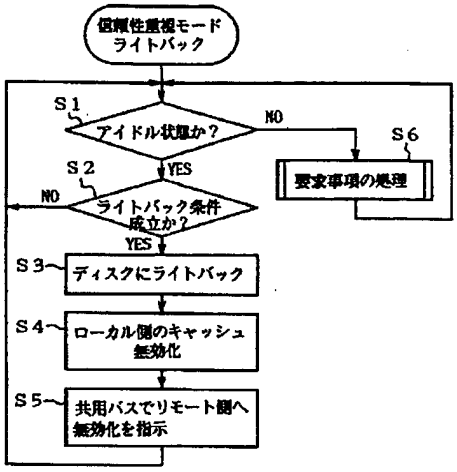
【図３】

図２のコントローラのハードウェア構成を示した実施例構成図



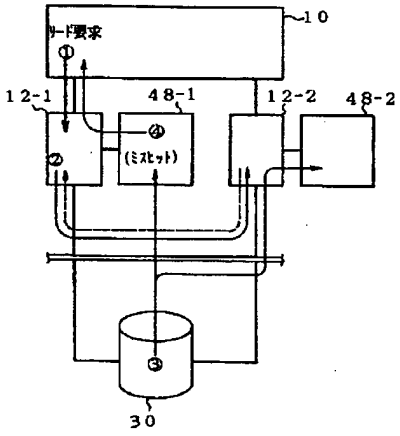
【図６】

図４の信頼性重視モードにおけるライトバック処理を示したフローチャート



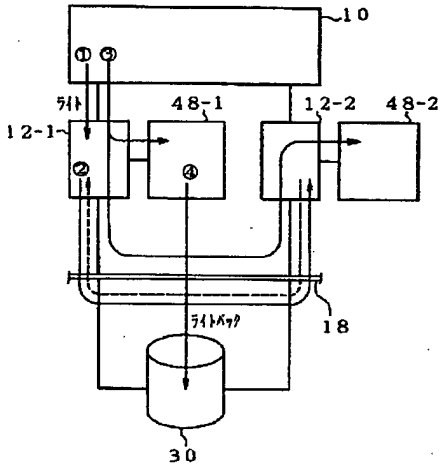
【図７】

信頼性重視モードでのリード要求に対しローカル側でミスヒットとなった場合の動作を示した説明図



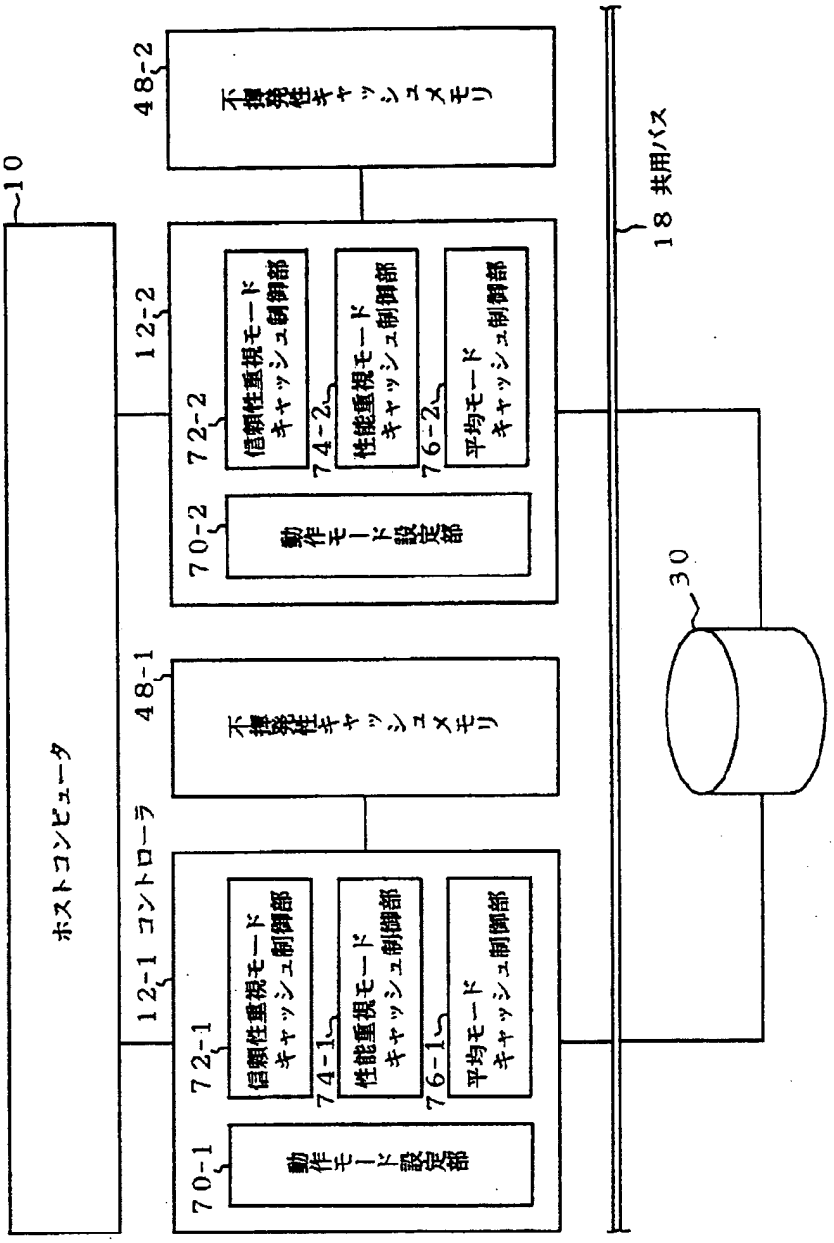
【図８】

信頼性重視モードでのライト要求に対するキャッシュ書込動作の説明図



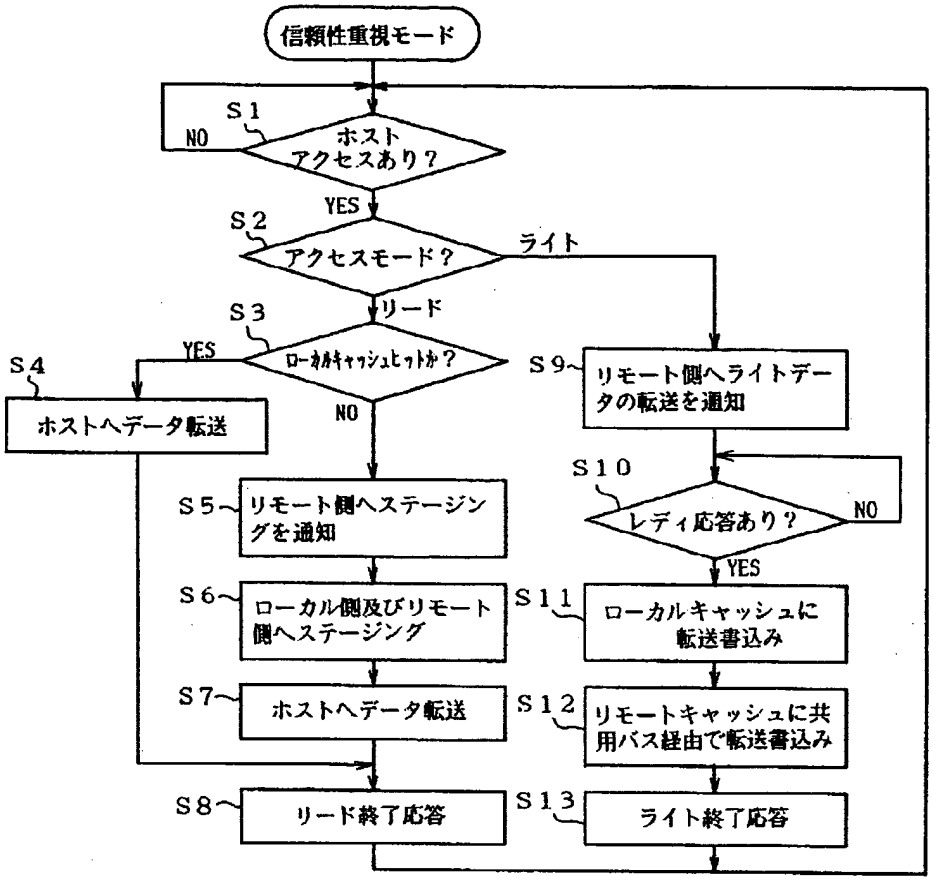
【図４】

本発明の第１発明の処理機能を示した説明図



【図５】

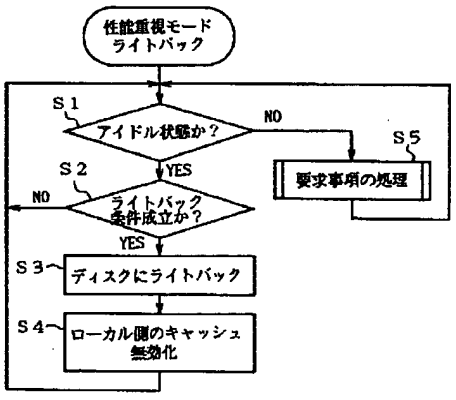
図４の信頼性重視モードの処理動作を示したフローチャート



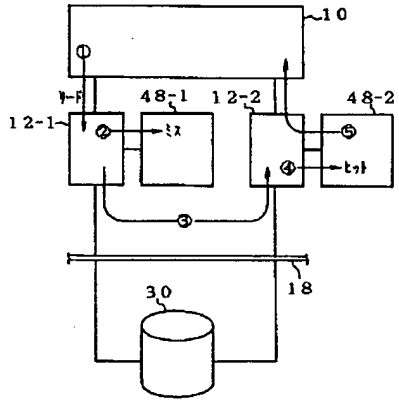
【図１０】

【図１１】

図４の性能重視モードにおけるライトバック処理を示したフローチャート

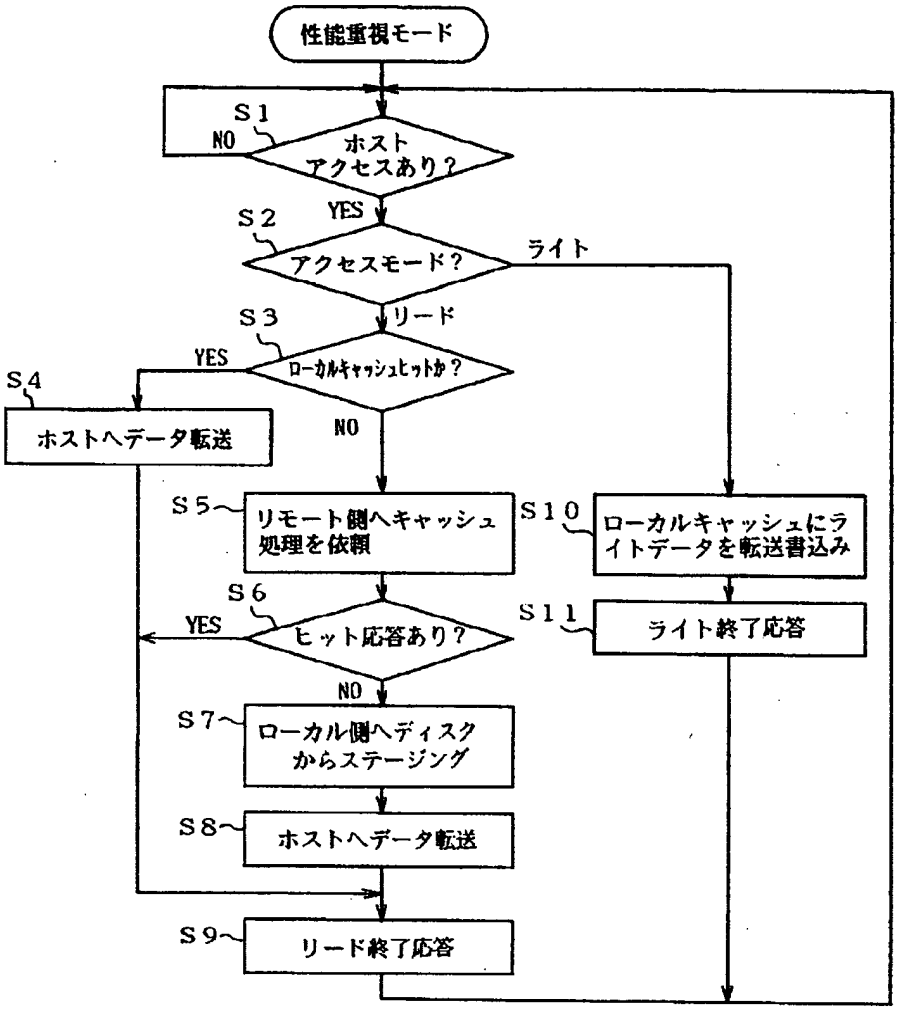


性能重視モードでのリード要求に対しローカル側でミスヒット、リモート側でヒットとなった場合の動作を示した説明図



【図９】

図４の性能重視モードの処理動作を示したフローチャート



【図２１】

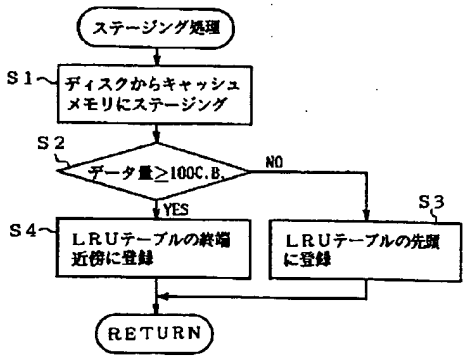
【図２９】

図１５の空スペース管理テーブルの説明図

64

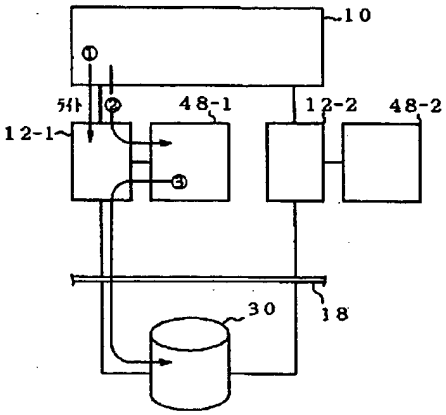
エントリ	先頭ブロック	内　　容
1	CB0203	キャッシュブロック１個分の空スペース有り
	CB0004	
2	CB0102	キャッシュブロック２個分の空スペース有り
3	CB00n-2	キャッシュブロック３個分の空スペース有り
⋮	⋮	⋮
N	CB00n-1	キャッシュブロックN個分の空スペース有り

大量シーケンシャルデータのステージング処理を示したフローチャート



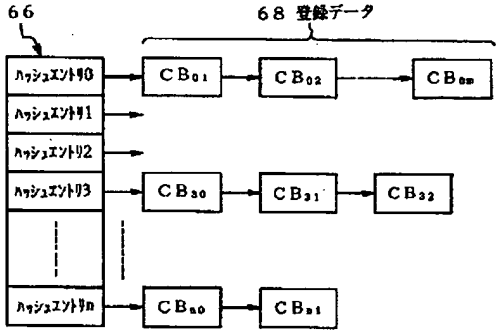
【図１２】

性能重視モードでのライト要求に対するキャッシュ書込動作の説明図



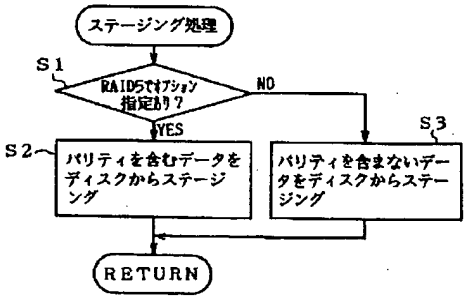
【図１８】

図１５のハッシュテーブルの説明図



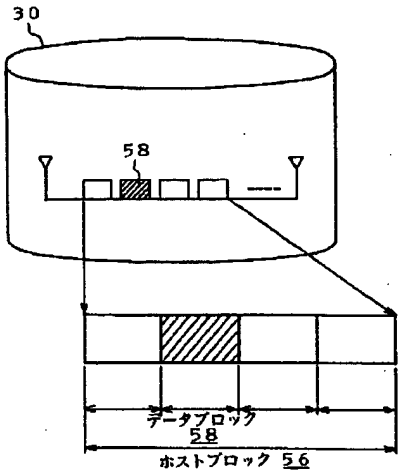
【図３０】

パリティを含むステージング処理を示したフローチャート



【図１７】

ホストブロックとデータブロックとの関係を示した説明図



【図２０】

図１５のキャッシュメモリの使用状態の一例を示した説明図

48

	00	01	02	03	04	05	06	07
00	R	R	W	R	R	R	R	R
01	R	W	W	W	W	W	W	W
02	R	空	空	R	R	R	R	R
03	R	空	R	R	R	R	R	R
04	空	R	R	R	R	R	R	R
n-3	R	R	R	R	W	W	W	W
n-2	空	空	空	R	R	R	R	R
n-1	空	空	空	空	空	空	空	空
n	空	空	空	空	空	空	空	空

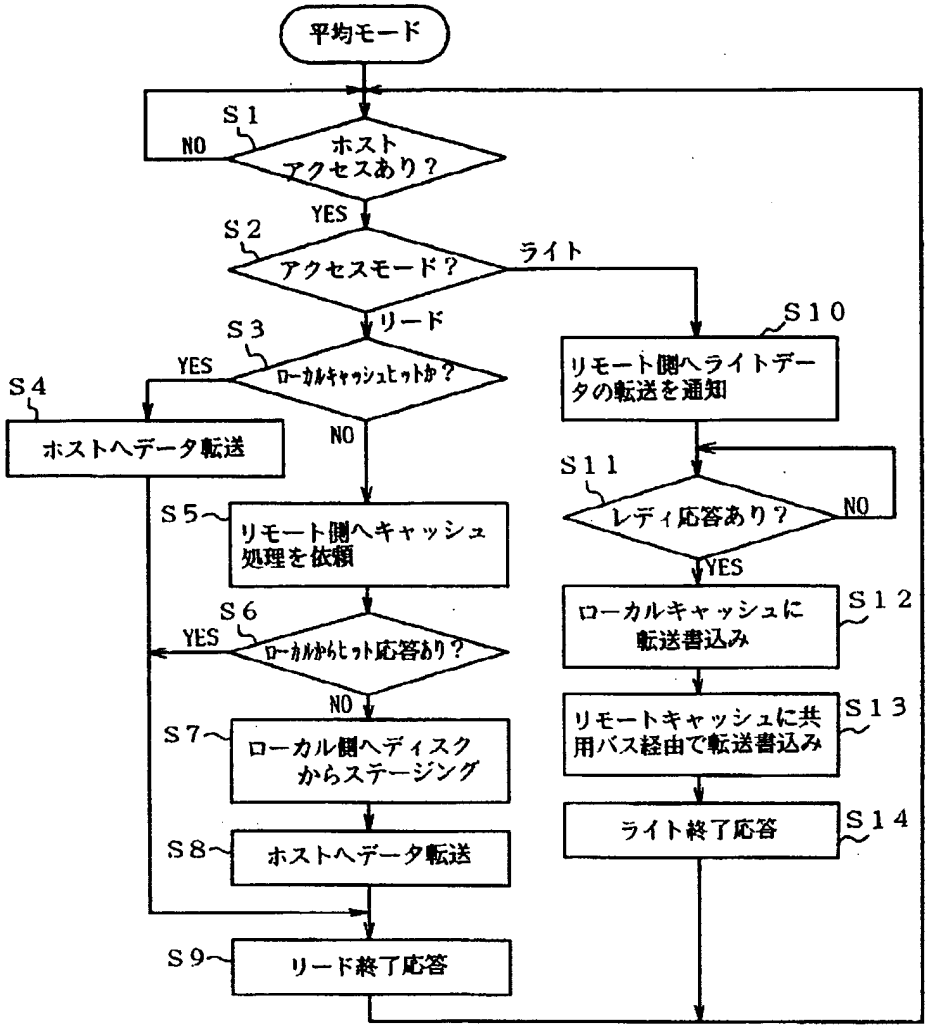
R：有効データの存在するキャッシュブロック

W：ダーティデータの存在するキャッシュブロック

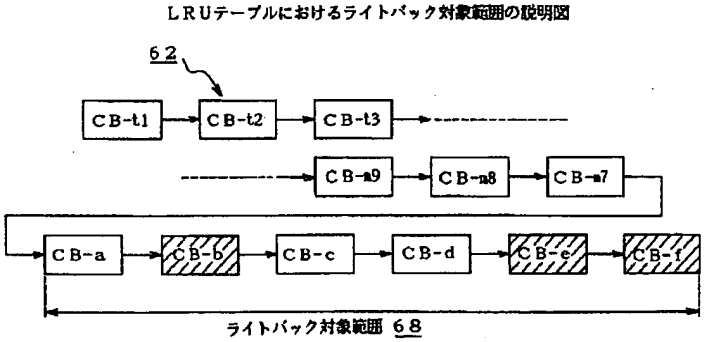
空：空キャッシュブロック

【図１３】

図４の平均モードの処理動作を示したフローチャート

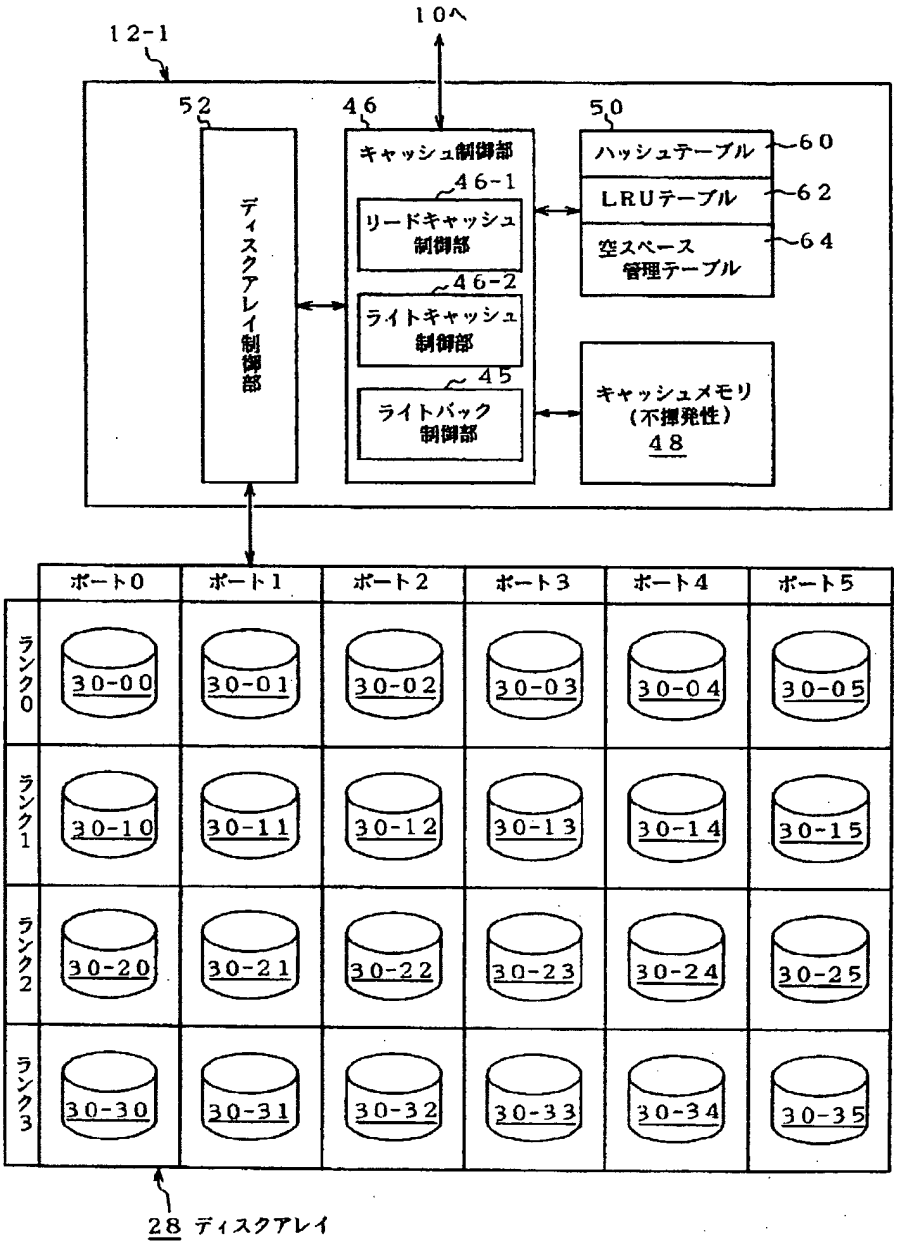


【図３２】



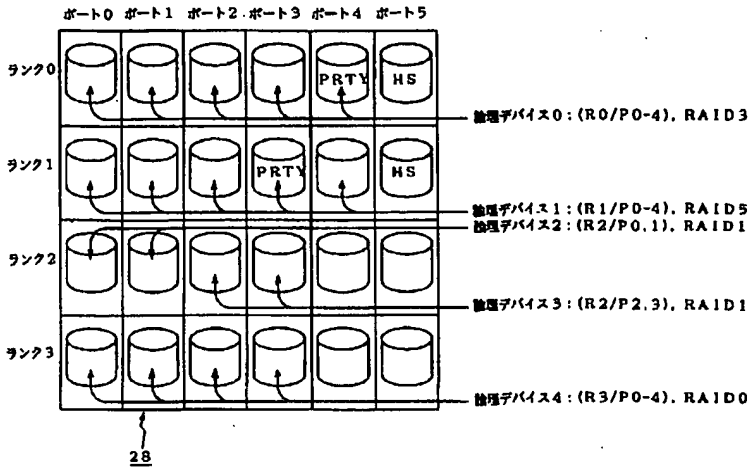
【図１５】

ディスクアレイを対象とした第２発明の処理機能を示した説明図



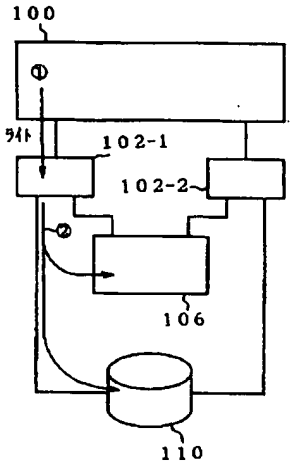
【図２２】

図１５のディスクアレイに対するＲＡＩＤ０，１，３，５のディスク割当てを示す論理デバイスの説明図



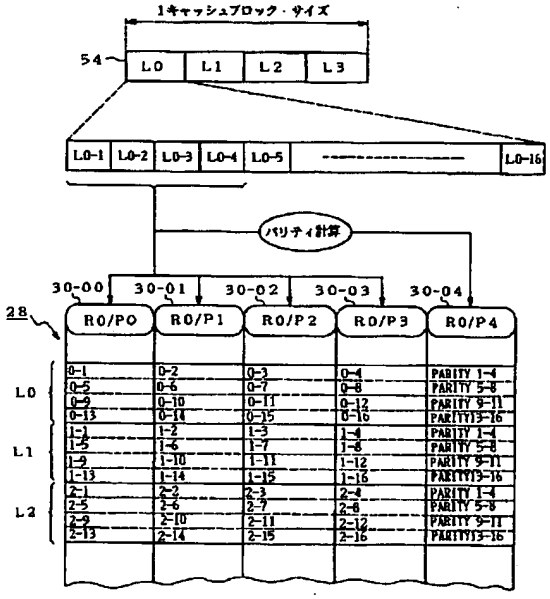
【図４４】

図４１の従来装置のライト処理を示した説明図



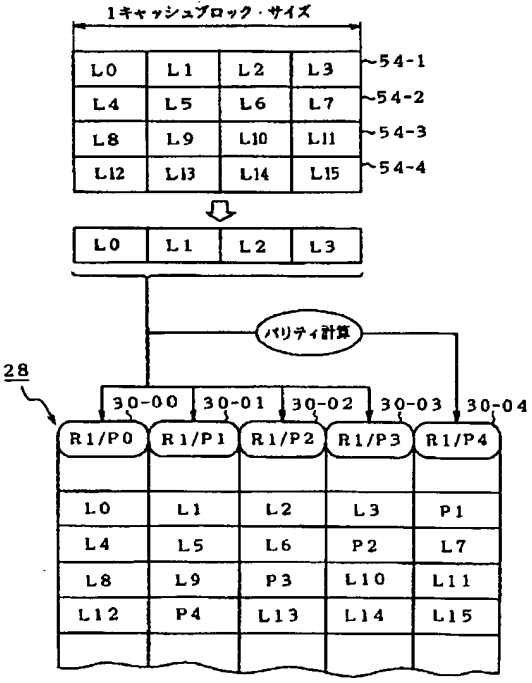
【図２３】

ＲＡＩＤ３におけるキャッシュブロック、ストライピング及びディスク格納状態を示した説明図



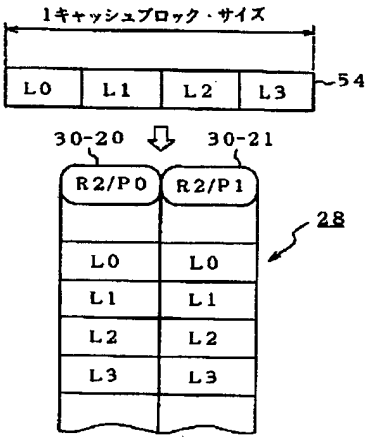
【図２４】

ＲＡＩＤ５におけるキャッシュブロック、ストライピング、ディスク格納状態を示した説明図



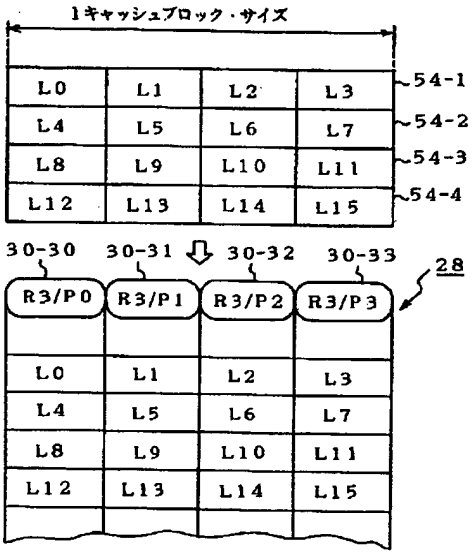
【図２５】

RAID1におけるキャッシュブロックとディスク格納状態を示した説明図



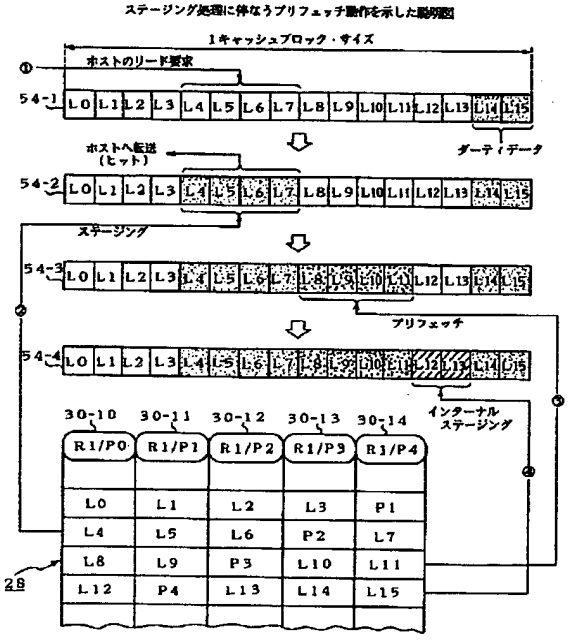
【図２６】

RAID0におけるキャッシュブロックとディスク格納状態を示した説明図



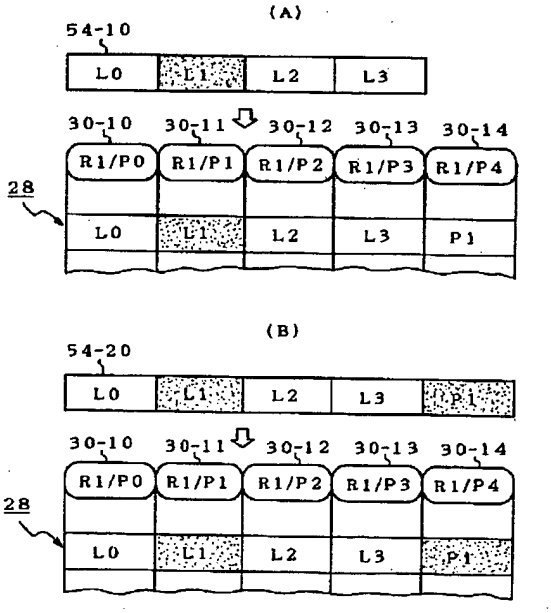
【図２７】

ステージング処理に伴なうプリフェッチ動作を示した説明図



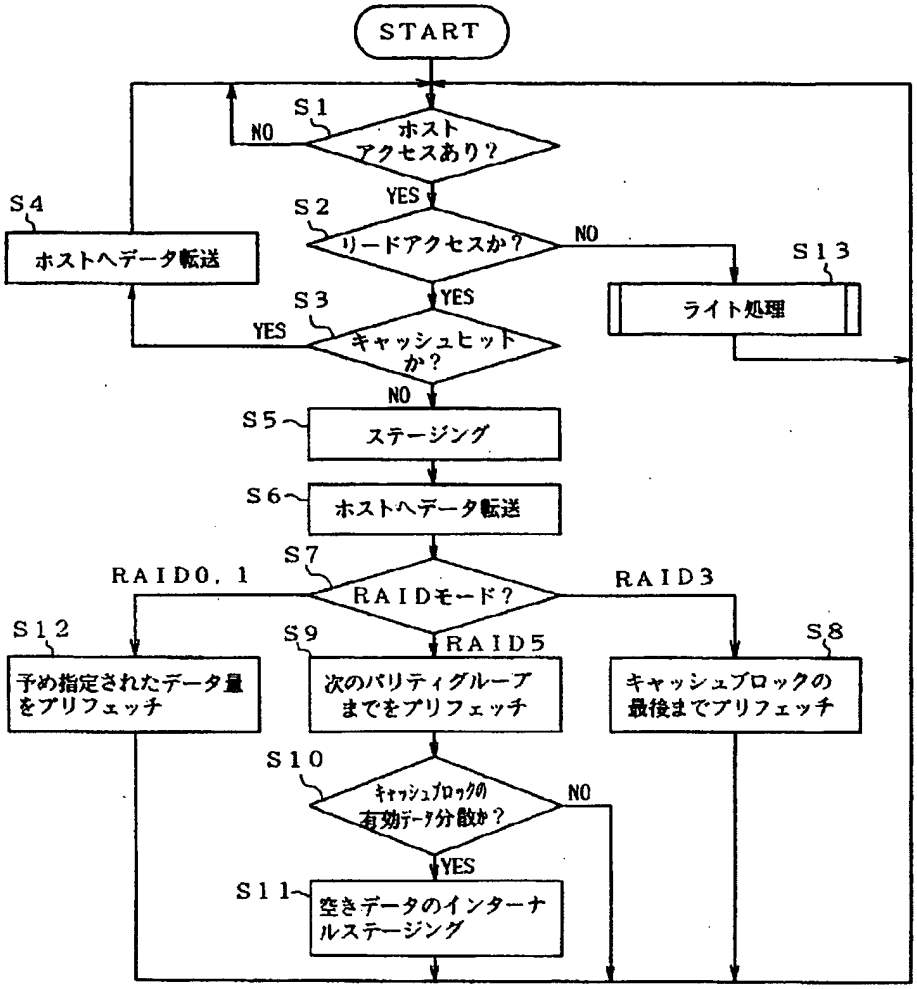
【図３１】

パリティを含まない場合と含む場合のステージングに対応するライトバック処理の相違を示した説明図



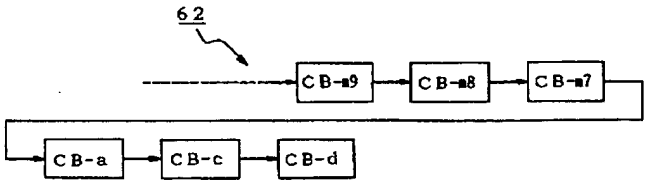
【図２８】

プリフェッチのステージングを含む図１５のリード処理を示したフローチャート



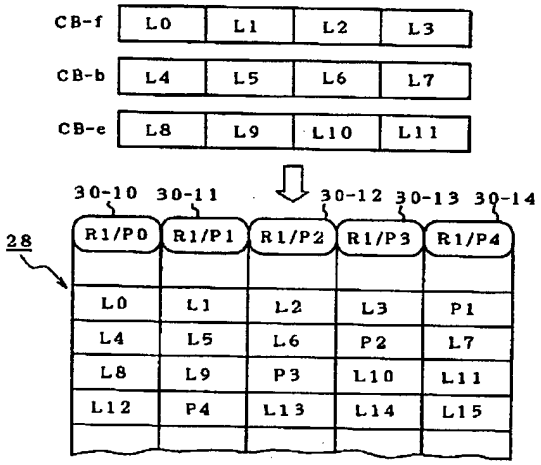
【図３４】

ライトバック完了後のＬＲＵテーブルの説明図



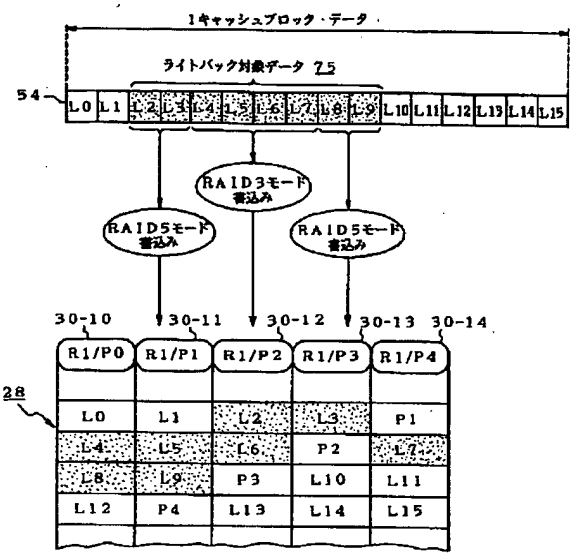
【図３３】

ライトバック対象範囲におけるキャッシュブロック並び替えとライトバック動作を示した説明図



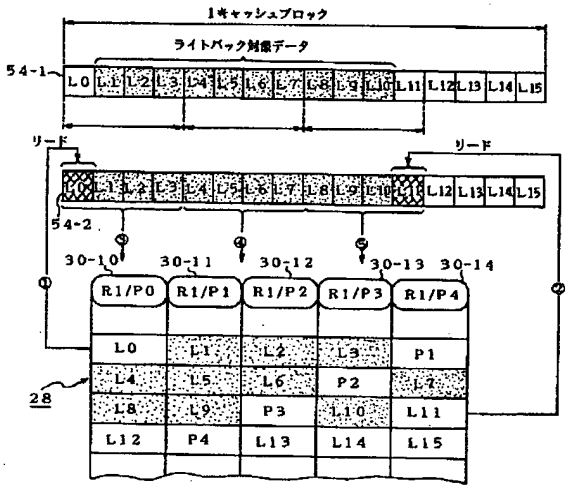
【図３６】

RAID5におけるライトバック動作の説明図



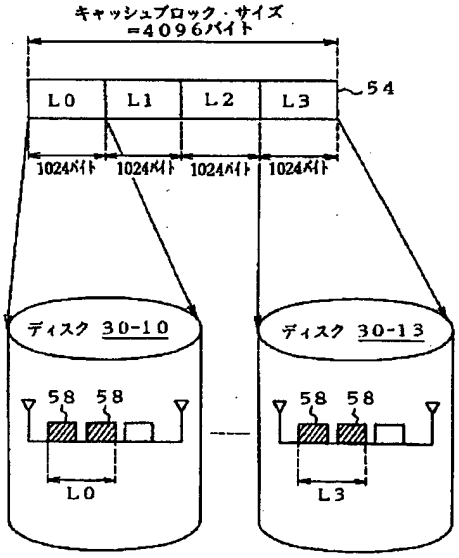
【図３７】

RAID5において不足ブロックをキャッシュブロックにステージングするライトバック動作の説明図



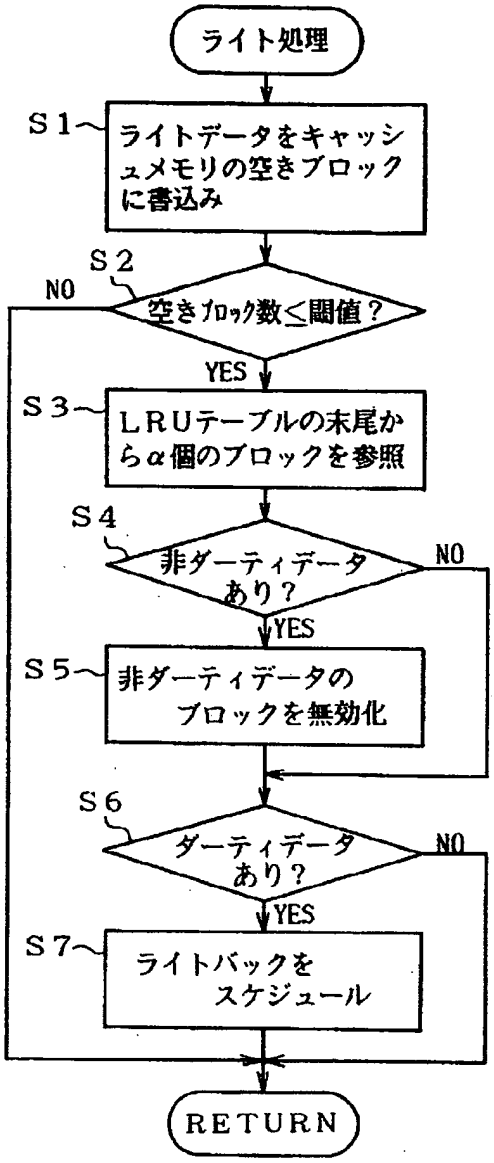
【図４０】

最適なキャッシュブロックサイズの決め方の一例を示した説明図



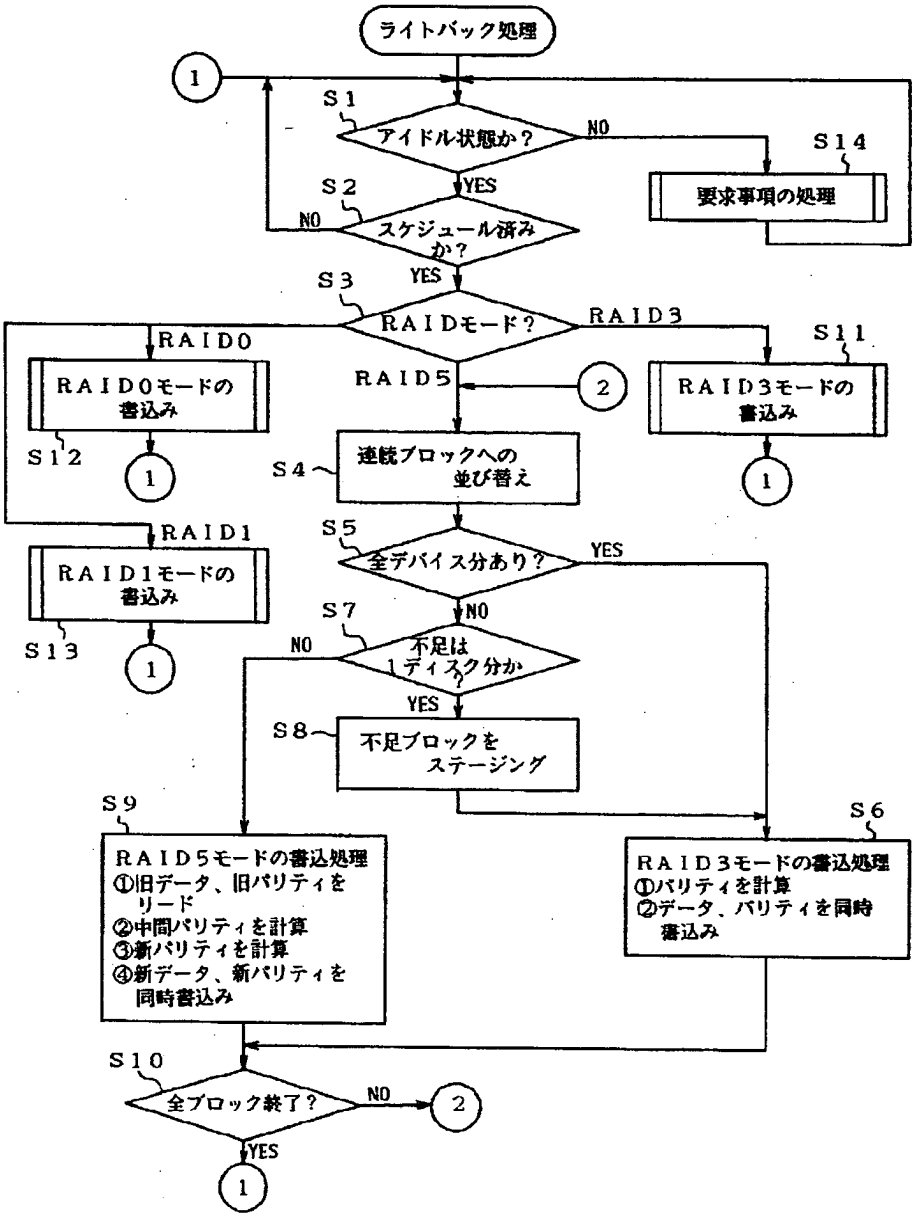
【図３５】

図２７のライト処理の詳細を示したフローチャート



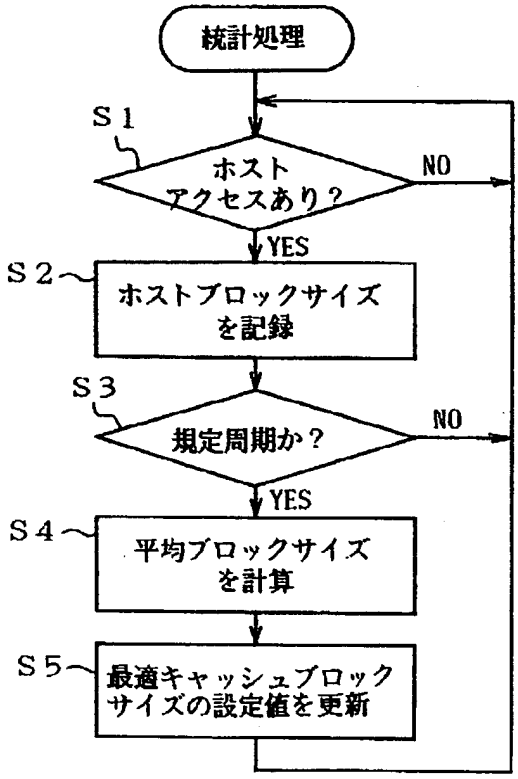
【図３８】

ライト完了後に行うライトバック処理を示したフローチャート



【図３９】

キャッシュブロックサイズの求める統計処理のフローチャート



【図４２】

【図４３】

図４１の従来装置のキャッシュヒット時のリード処理を示した説明図

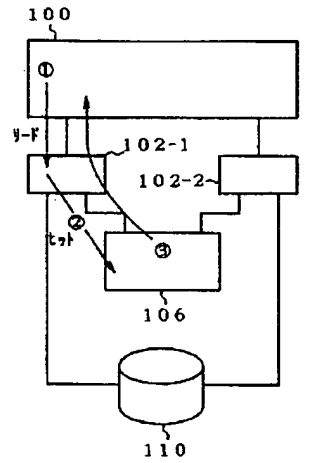
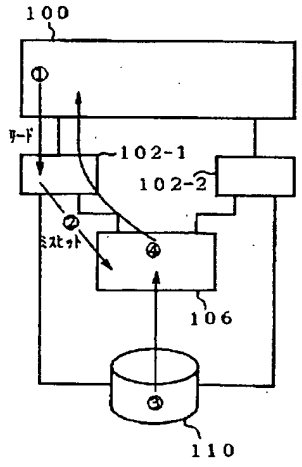
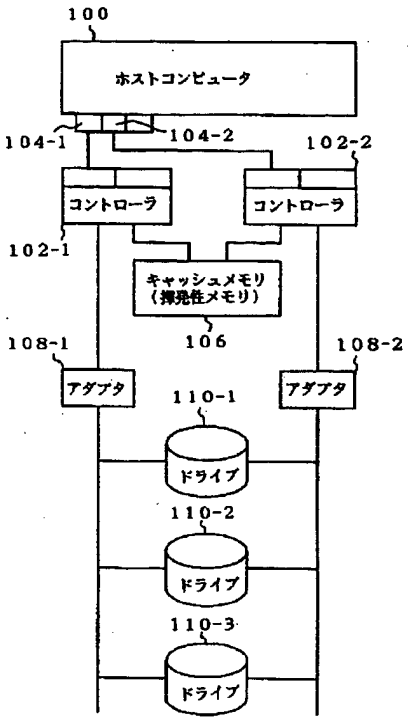


図４１の従来装置のキャッシュミスヒット時のリード処理を示した説明図



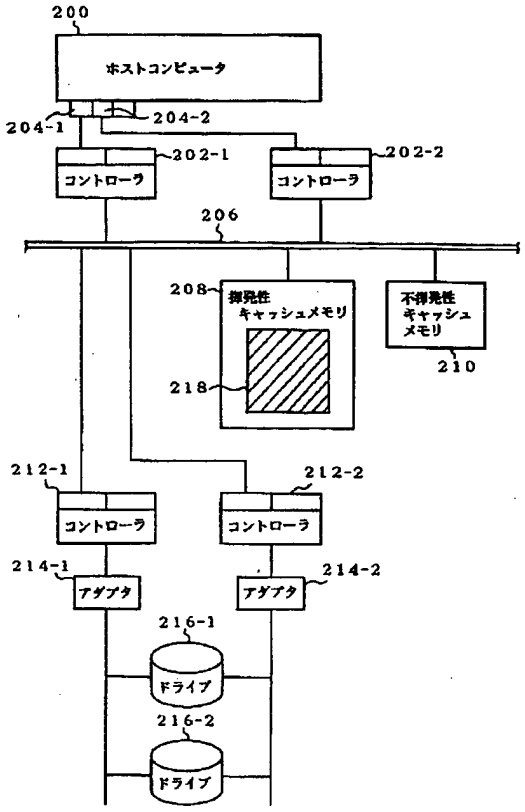
【図４１】

揮発性メモリを用いた従来のディスクキャッシュ装置の説明図



【図４５】

揮発性メモリと不揮発性メモリを混在させた従来装置の説明図



【図４６】

【図４７】

図４５の従来装置のキャッシュミスヒット時のリード処理を示した説明図

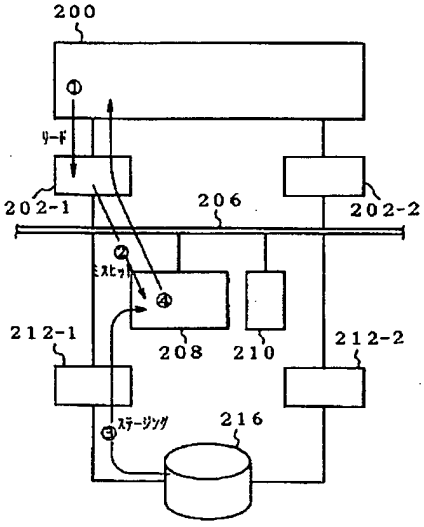


図４５の従来装置のライト処理を示した説明図